GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:16 ; Search time 31 Seconds
(without alignments)
34.387 Million cell updates/sec

Title: US-09-719-494-10
Perfect score: 36
Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	100.0	8	16	AAR83942	MHC class I restri
2	36	100.0	8	20	AAY13407	HSV specific antig
3	36	100.0	8	20	AAY10350	T cell epitope/MHC
4	36	100.0	8	21	AAB12542	gB-specific H-2b r
5	36	100.0	8	21	AAY67358	HSV glycoprotein B
6	36	100.0	19	23	AAE13452	Herpes simplex vir
7	36	100.0	19	23	AAE13453	Herpes simplex vir
8	36	100.0	22	23	AAU82071	Antigenic peptide
9	36	100.0	23	20	AAY13421	Conjugated control
10	36	100.0	26	20	AAY13420	Conjugated peptide
11	36	100.0	27	20	AAY13419	Conjugated peptide
12	36	100.0	30	23	AAE13454	Herpes simplex vir
13	36	100.0	38	23	AAU82093	T-cell specific bi
14	36	100.0	40	23	AAU82094	T-cell specific bi
15	36	100.0	694	12	AAR14666	Truncated HSVgB po
16	36	100.0	795	19	AAW72062	HSV-2 strain SB5 C
17	36	100.0	854	19	AAW72113	HSV-2 strain SB5 C
18	36	100.0	903	6	AAP50312	Herpes simplex vir
19	36	100.0	903	7	AAP60244	Herpes simplex vir
20	36	100.0	903	8	AAP70426	Recombinant herpes
21	36	100.0	903	8	AAP71135	Herpes Simplex Vir
22	36	100.0	904	9	AAP80914	Sequence of Herpes
23	36	100.0	904	12	AAR14665	HSVgB polypeptide.
24	36	100.0	904	14	AAR41778	Glycoprotein B (gB
25	36	100.0	904	14	AAR41779	Glycoprotein B (gB
26	36	100.0	904	17	AAW00375	HSV-1 glycoprotein
27	36	100.0	904	17	AAW00376	HSV-2 glycoprotein
28	36	100.0	904	18	AAW34552	Herpes simplex vir
29	36	100.0	904	18	AAW34553	Herpes simplex vir
30	36	100.0	904	19	AAW72193	HSV-2 strain SB5 C
31	36	100.0	904	22	AAB74441	Herpes simplex vir
32	36	100.0	904	22	AAB74442	Herpes simplex vir
33	36	100.0	904	23	AAE17812	Herpes simplex vir
34	36	100.0	905	9	AAP80915	Sequence of Herpes
35	36	100.0	907	8	AAP71136	Herpes Simplex Vir
36	36	100.0	973	8	AAP70769	Glycoprotein B of
37	36	100.0	973	12	AAR14680	HSV surface antige
38	35	97.2	885	17	AAR92747	SA8 virus gB glyco
39	35	97.2	891	17	AAR92746	B virus gB glycopr
40	35	97.2	891	19	AAW70293	Simian herpesvirus
41	32	88.9	8	21	AAY67359	HSV glycoprotein B
42	30	83.3	344	23	AAU76345	Yeast D-arabinose
43	30	83.3	372	21	AAB52570	Helicobacter pylor
44	30	83.3	465	18	AAW55734	H. pylori ORF 07ee
45	30	83.3	586	22	AAE00651	Aspergillus nidula

ALIGNMENTS

RESULT 1
AAR83942
ID AAR83942 standard; peptide; 8 AA.
XX
AC AAR83942;
XX
DT 05-JUN-1996 (first entry)
XX
DE MHC class I restricted antigenic peptide #12.
XX
KW MHC class I; antigen; MAGE; melanoma; breast cancer; bladder cancer;
KW Titermax; cytotoxic T-lymphocyte; tumour; pathogenic disease; bacteria;
KW parasite; human; animal.
XX
OS Synthetic.
XX
PN WO9528958-A1.
XX
PD 02-NOV-1995.
XX
PF 21-APR-1995; 95WO-US04975.
XX
PR 22-APR-1994; 94US-0233496.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PI Dyall R, Nikolic-Zugic J;
XX
DR WPI; 1995-382848/49.
XX
PT Cytotoxic T-cell induction by MHC class I-restricted peptide in
PT adjuvant - useful for treating tumours and bacterial or parasitic
PT pathogenic diseases

XX
PS Claim 11; Page 38; 50pp; English.
XX
CC The sequences given in AAR83931-49 are MHC class I restricted 8-12 amino
CC acid antigenic peptides. This peptide is derived from Herpes simplex
CC virus gB protein. These peptides may be administered to a subject
CC in combination with a suitable adjuvant, pref. Titermax (RTM), to
CC induce cytotoxic T-lymphocytes. This method may be used in the
CC treatment of a tumour or a pathogenic disease, esp. diseases of
CC bacterial or parasitic origin, in humans and animals, e.g monkeys,
CC dogs, cows, horses, etc.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 36; DB 16; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 1 SSIEFARL 8

RESULT 2
AAY13407
ID AAY13407 standard; peptide; 8 AA.
XX
AC AAY13407;
XX
DT 19-JUL-1999 (first entry)
XX
DE HSV specific antigenic peptide B1 of glycoprotein B.
XX
KW Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;
KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;
KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;
KW immune response; genetic immunisation.
XX
OS Herpes simplex virus type 1.
XX
PN WO9916710-A1.
XX
PD 08-APR-1999.
XX
PF 29-SEP-1998; 98WO-US20681.
XX
PR 30-SEP-1997; 97US-0060422.
XX
PA (CELS-) CEL-SCI CORP.
PA (UYNE-) UNIV NORTHEASTERN OHIO.
XX
PI Rosenthal KS, Zimmerman DH;
XX
DR WPI; 1999-312418/26.
XX
PT New immunogenic conjugated polypeptides
XX
PS Claim 2; Page 44; 69pp; English.
XX
CC The invention provides immunogenic conjugated polypeptides which comprise
CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory
CC peptide and which promote binding to T cells. The novel immunogenic
CC conjugated polypeptides are effective as an immunogen in a vaccine for
CC treatment or prevention of infection by HSV and are represented by the
CC formula P1-x-P2 or P2-x-P1; where P1 = a HSV specific antigenic peptide
CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein
CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and
CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of
CC an immunoprotein which promotes binding to a class of subclass of T cells
CC and which directs a predominantly TH1 type immune response to the peptide
CC P1; and x = a covalent bond or a cleavable or non-cleavable peptide
CC linking group. The conjugated polypeptides can elicit an immune response
CC for neutralizing HSV and killing HSV infected cells. They can be used for
CC the treatment or prevention of HSV infection. In addition, DNA encoding
CC the polypeptide can be used for genetic immunization. They can also be
CC used to diagnose the presence of infection, active or latent, in an
CC individual by HSV by mixing T cells from the individual with the
CC polypeptide and detecting a reaction between the T cells and the
CC polypeptide.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 36; DB 20; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 1 SSIEFARL 8

RESULT 3
AAY10350
ID AAY10350 standard; Peptide; 8 AA.
XX
AC AAY10350;
XX
DT 12-MAY-1999 (first entry)
XX
DE T cell epitope/MHC ligand SEQ ID NO:280.
XX
KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW immunisation; tumour; infectious disease; immunotherapy; cancer;
KW malignant melanoma; viral disease; hepatitis; AIDS.
XX
OS Synthetic.
OS Herpes simplex virus.
XX
PN WO9902183-A2.
XX
PD 21-JAN-1999.
XX
PF 10-JUL-1998; 98WO-US14289.
XX
PR 10-DEC-1997; 97US-0988320.
PR 10-JUL-1997; 97CA-2209815.
XX
PA (CTLI-) CTL IMMUNOTHERAPIES CORP.
XX
PI Kuendig TM, Simard JJL;
XX
DR WPI; 1999-120514/10.
XX
PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT of antigen in the lymphatic system of a mammal so as to provide a
PT sustained CTL response, used to treat, e.g. AIDS
XX
PS Disclosure; Page 35; 199pp; English.
XX
CC The present invention describes a method of inducing and/or sustaining
CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC method comprises: (a) delivering an antigen to the mammal at a level to
CC induce an immunological CTL response in the mammal; and (b) maintaining
CC the level of the antigen in the mammal's lymphatic system to maintain
CC the immunologic CTL response. The method can be used for the delivery of
CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC gene antigen, or a viral antigen. They can be used for the treatment of
CC disease such as cancer, e.g. malignant melanoma or infectious disease,
CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC to the lymphatic system provides for potent CTL stimulation that takes
CC place in the milieu of the lymphoid organ, and it sustains stimulation
CC that is necessary to keep CTL active, cytotoxic and recirculating
CC through the body. AAY10071 to AAY10639 represent examples of peptide
CC antigens given in the present invention.
XX

```
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 36;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db       1 SSIEFARL 8


RESULT 4
AAB12542
ID   AAB12542 standard; Peptide; 8 AA.
XX
AC   AAB12542;
XX
DT   03-NOV-2000  (first entry)
XX
DE   gB-specific H-2b restricted peptide SEQ ID NO:2.
XX
KW   Herpes virus; HSV; antibacterial; antiviral; antiparasitic; nootropic;
KW   antidiabetic; neuroprotective; antiparkinsonian; vaccine; gene therapy;
KW   amplicon; packaging; delivery.
XX
OS   Synthetic.
XX
PN   WO200034497-A2.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;   99WO-US29120.
XX
PR   09-DEC-1998;   98US-0111630.
PR   26-OCT-1999;   99US-0161374.
XX
PA   (BREA/) BREAKEFIELD X O.
PA   (CHIO/) CHIOCCA E A.
PA   (SAEK/) SAEKI Y.
PA   (FRAE/) FRAEFEL C.
PA   (TOBL/) TOBLER K.
PA   (ACKE/) ACKERMANN M.
PA   (SUTE/) SUTER M.
PA   (ADEM/) ADEMA G J.
PA   (SHOR/) SHORTMAN K.
XX
PI   Breakefield XO,  Chiocca EA,  Saeki Y,  Fraefel C,  Tobler K;
PI   Ackermann M,  Suter M,  Adema GJ,  Shortman K;
XX
DR   WPI; 2000-431309/37.
XX
PT   Herpes virus amplicon packaging system comprising herpes virus amplicon
PT   vector and packaging vector having single cloning vehicle containing
PT   replication proficient but packaging defective entire herpes virus
PT   genome -
XX
PS   Example 3; Page 49; 78pp; English.
XX
CC   The present invention describes a Herpes virus (HSV) amplicon packaging
CC   system (I) comprising a HSV amplicon vector (II) and a packaging vector
CC   (III) comprising a single cloning vehicle which contains a replication
CC   proficient and packaging defective entire HSV genome. (II) is packaged
CC   into infectious particles that are substantially (III) of helper virus
CC   contamination by co-transfection with (III) which provides helper virus
CC   function. (I) is used for generating recombinant HSV vectors which are
CC   used for inducing protective immunity against herpes simplex virus
CC   (HSV-1) and any viral, bacterial or parasitic pathogen in an animal.
CC   The defective HSV-1 genome packaging system can be used to package a
CC   wide range of desired nucleotide segments, preferably a DNA segment,
CC   into an empty HSV particle. The HSV particles can be used to deliver
CC   heterologous DNA to a target cell. The heterologous sequence can encode
CC   any desired protein such as a therapeutic protein e.g. one that
CC   compensates for an inherited or acquired deficiency, e.g. tyrosine
CC   hydroxylase for the treatment of Parkinson's disease, neurotrophic
CC   factors including neurotrophins, for the treatment of Alzheimer's
CC   disease, nerve growth factor receptor and the trk receptor, hypoxanthine-
CC   guanine phosphoribosyl transferase (HGPRT) for the treatment of Lesch
CC   Nyhan disease, beta-hexosaminidase alpha chain for the treatment of Tay
CC   Sachs disease, insulin for the treatment of diabetes. The present
CC   sequence represents a peptide used in an example from the present
CC   invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 36;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db       1 SSIEFARL 8


RESULT 5
AAY67358
ID   AAY67358 standard; peptide; 8 AA.
XX
AC   AAY67358;
XX
DT   25-APR-2000  (first entry)
XX
DE   HSV glycoprotein B, SSI peptide sequence used as a target peptide.
XX
KW   Target peptide; cytotoxic T lymphocyte; CTL; CTL immune response;
KW   cellular immune response induction method; vaccine; human; tumour;
KW   glycoprotein B.
XX
OS   Herpes simplex virus.
XX
PN   WO9963945-A2.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US13146.
XX
PR   12-JUN-1998;   98US-0089055.
PR   30-OCT-1998;   98US-0106339.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Nikolic-Zugic J,  Dyall R,  Houghton AN;
XX
DR   WPI; 2000-126432/11.
XX
PT   Induction of a cellular immune response to a weakly immunogenic
PT   protein, used to target and kill tumour cells -
XX
PS   Claim 16; Page 24; 44pp; English.
XX
CC   This sequence represents a Herpes simplex virus glycoprotein B peptide
CC   used as a target peptide in the method of the invention. The invention
CC   relates to a method for inducing a cytotoxic T lymphocyte (CTL) immune
CC   response to non/weakly-immunogenic proteins which are expressed on tumour
CC   cells. The method for inducing a cellular immune response to a
CC   non-immunogenic or weakly immunogenic target peptide expressed on tumour
CC   cells of a mammalian subject comprises administering antigen to induce a
CC   cellular immune response to the target peptide. The antigen comprises an
CC   immunogenic portion having a major histocompatibility complex (MHC)
CC   binding domain which binds to the MHC and an immune recognition domain
CC   which is recognized by T-cells. The antigen is derived from the target
CC   peptide such that the MHC-binding portion binds to MHC with a greater
CC   affinity than the target peptide without material alteration of the
CC   immune recognition portion. The methods are used for inducing a cellular
CC   immune response to a non-immunogenic or weakly immunogenic target peptide
CC   expressed on tumour cells of a mammalian subject. The antigens and
CC   immunogens of the invention, as well as polynucleotides encoding them,
```

```
CC   are used in vaccine compositions against tumour cells.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 36;  DB 21;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db      1 SSIEFARL 8


RESULT 6
AAE13452
ID   AAE13452 standard; peptide; 19 AA.
XX
AC   AAE13452;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Herpes simplex virus MHC class I peptide antigen #4.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatability complex; MHC; therapy; immune response;
KW   malignancy.
XX
OS   Herpes simplex virus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          12..19
FT                   /note= "MHC class I epitope"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatability complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is Herpes simplex virus MHC class I peptide antigen.
XX
SQ   Sequence   19 AA;

  Query Match            100.0%;  Score 36;  DB 23;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 0.11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db     12 SSIEFARL 19


RESULT 7
AAE13453
ID   AAE13453 standard; peptide; 19 AA.
XX
AC   AAE13453;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Herpes simplex virus MHC class I peptide antigen #5.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatability complex; MHC; therapy; immune response;
KW   malignancy.
XX
OS   Herpes simplex virus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "MHC class I epitope"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          12..19
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatability complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is Herpes simplex virus MHC class I peptide antigen.
XX
SQ   Sequence   19 AA;

  Query Match            100.0%;  Score 36;  DB 23;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 0.11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db      1 SSIEFARL 8
```

```
RESULT 8
AAU82071
ID   AAU82071 standard; peptide; 22 AA.
XX
AC   AAU82071;
XX
DT   09-APR-2002  (first entry)
XX
DE   Antigenic peptide extgB1 associated with Herpes Simplex Virus.
XX
KW   T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;
KW   peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;
KW   immunological disorder; immune response; human immunodeficiency virus;
KW   herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;
KW   acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;
KW   autoimmune myocarditis; cytostatic; antiinflammatory.
XX
OS   Synthetic.
XX
PN   WO200189286-A2.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-US16793.
XX
PR   24-MAY-2000; 2000US-206548P.
XX
PA   (CELS-) CEL-SCI CORP.
XX
PI   Zimmerman DS,  Sarin PS;
XX
DR   WPI; 2002-083037/11.
XX
PT   New T cell binding ligand peptide for treating immunological disorders
PT   such as herpes simplex virus, tuberculosis, cancers, acquired
PT   immunodeficiency syndrome and allergies  -
XX
PS   Disclosure; Page 27; 110pp; English.
XX
CC   The present invention relates to novel T-cell binding ligand (TCBL)
CC   peptides (e.g. peptide G' (modified human MHC class II beta chain
CC   peptide G, peptide J (human beta-2-microglobulin peptide) and HIV-1
CC   peptides) and TCBL peptide constructs for treating immunological
CC   disorders. The peptide constructs are useful for eliciting a cellular
CC   immune response in a human patient. The method comprises administering
CC   the peptide construct to the patient preferably in combination with an
CC   immune response adjuvant. The peptide constructs in the form of
CC   conjugated peptides are useful for eliciting a cellular immune response
CC   in a patient exposed to or at risk for exposure to the human
CC   immunodeficiency virus (HIV). The TCBL peptides are useful for treating a
CC   patient suffering from an immunological disorder such as herpes simplex
CC   virus (HSV) infection, malaria, tuberculosis, cancers, acquired
CC   immunodeficiency syndrome (AIDS), allergies, autoimmune diseases
CC   (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune
CC   myocarditis, diabetes and lupus) by administering a peptide construct
CC   comprising a TCBL peptide bonded to an antigenic peptide associated with
CC   the disorder. Unlike prior art peptide conjugates, a modified version of
CC   peptide G has long range stabilisation and also enhances the immune
CC   response. AAU82019-AAU82114 represent T-cell specific binding ligand
CC   peptides, peptide constructs or peptides used in their construction.
XX
SQ   Sequence   22 AA;

  Query Match             100.0%;  Score 36;  DB 23;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 0.12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db       7 SSIEFARL 14
```

```
RESULT 9
AAY13421
ID   AAY13421 standard; peptide; 23 AA.
XX
AC   AAY13421;
XX
DT   19-JUL-1999  (first entry)
XX
DE   Conjugated control peptide L-B1.
XX
KW   Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;
KW   immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;
KW   ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;
KW   immune response; genetic immunisation.
XX
OS   Synthetic.
XX
PN   WO9916710-A1.
XX
PD   08-APR-1999.
XX
PF   29-SEP-1998;   98WO-US20681.
XX
PR   30-SEP-1997;   97US-0060422.
XX
PA   (CELS-) CEL-SCI CORP.
PA   (UYNE-) UNIV NORTHEASTERN OHIO.
XX
PI   Rosenthal KS,  Zimmerman DH;
XX
DR   WPI; 1999-312418/26.
XX
PT   New immunogenic conjugated polypeptides
XX
PS   Example 1; Page 33; 69pp; English.
XX
CC   The invention provides immunogenic conjugated polypeptides which comprise
CC   a herpes simplex virus peptide (HSV) linked to an immunomodulatory
CC   peptide and which promote binding to T cells. The novel immunogenic
CC   conjugated polypeptides are effective as an immunogen in a vaccine for
CC   treatment or prevention of infection by HSV and are represented by the
CC   formula P1-x-P2 or P2-x-P1; where P1 = a HSV specific antigenic peptide
CC   from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein
CC   B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and
CC   glycoprotein F; P2 = an immunomodulatory peptide which is a portion of
CC   an immunoprotein which promotes binding to a class of subclass of T cells
CC   and which directs a predominantly TH1 type immune response to the peptide
CC   P1; and x = a covalent bond or a cleavable or non-cleavable peptide
CC   linking group. The conjugated polypeptides can elicit an immune response
CC   for neutralizing HSV and killing HSV infected cells. They can be used for
CC   the treatment or prevention of HSV infection. In addition, DNA encoding
CC   the polypeptide can be used for genetic immunization. They can also be
CC   used to diagnose the presence of infection, active or latent, in an
CC   individual by HSV by mixing T cells from the individual with the
CC   polypeptide and detecting a reaction between the T cells and the
CC   polypeptide.
XX
SQ   Sequence   23 AA;

  Query Match             100.0%;  Score 36;  DB 20;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db      16 SSIEFARL 23
```

```
RESULT 10
AAY13420
ID   AAY13420 standard; peptide; 26 AA.
XX
AC   AAY13420;
```

XX
DT 19-JUL-1999 (first entry)
XX
DE Conjugated peptide G-B1.
XX
KW Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;
KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;
KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;
KW immune response; genetic immunisation.
XX
OS Synthetic.
XX
PN WO9916710-A1.
XX
PD 08-APR-1999.
XX
PF 29-SEP-1998; 98WO-US20681.
XX
PR 30-SEP-1997; 97US-0060422.
XX
PA (CELS-) CEL-SCI CORP.
PA (UYNE-) UNIV NORTHEASTERN OHIO.
XX
PI Rosenthal KS, Zimmerman DH;
XX
DR WPI; 1999-312418/26.
XX
PT New immunogenic conjugated polypeptides
XX
PS Claim 4; Page 45; 69pp; English.
XX
CC The invention provides immunogenic conjugated polypeptides which comprise
CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory
CC peptide and which promote binding to T cells. The novel immunogenic
CC conjugated polypeptides are effective as an immunogen in a vaccine for
CC treatment or prevention of infection by HSV and are represented by the
CC formula P1-x-P2 or P2-x-P1; where P1 = a HSV specific antigenic peptide
CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein
CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and
CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of
CC an immunoprotein which promotes binding to a class of subclass of T cells
CC and which directs a predominantly TH1 type immune response to the peptide
CC P1; and x = a covalent bond or a cleavable or non-cleavable peptide
CC linking group. The conjugated polypeptides can elicit an immune response
CC for neutralizing HSV and killing HSV infected cells. They can be used for
CC the treatment or prevention of HSV infection. In addition, DNA encoding
CC the polypeptide can be used for genetic immunization. They can also be
CC used to diagnose the presence of infection, active or latent, in an
CC individual by HSV by mixing T cells from the individual with the
CC polypeptide and detecting a reaction between the T cells and the
CC polypeptide.
XX
SQ Sequence 26 AA;

Query Match 100.0%; Score 36; DB 20; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 19 SSIEFARL 26

RESULT 11
AAY13419
ID AAY13419 standard; peptide; 27 AA.
XX
AC AAY13419;
XX
DT 19-JUL-1999 (first entry)
XX
DE Conjugated peptide J-B1.
XX
KW Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;
KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;
KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;
KW immune response; genetic immunisation.
XX
OS Synthetic.
OS Herpes simplex virus type 1.
XX
PN WO9916710-A1.
XX
PD 08-APR-1999.
XX
PF 29-SEP-1998; 98WO-US20681.
XX
PR 30-SEP-1997; 97US-0060422.
XX
PA (CELS-) CEL-SCI CORP.
PA (UYNE-) UNIV NORTHEASTERN OHIO.
XX
PI Rosenthal KS, Zimmerman DH;
XX
DR WPI; 1999-312418/26.
XX
PT New immunogenic conjugated polypeptides
XX
PS Claim 4; Page 45; 69pp; English.
XX
CC The invention provides immunogenic conjugated polypeptides which comprise
CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory
CC peptide and which promote binding to T cells. The novel immunogenic
CC conjugated polypeptides are effective as an immunogen in a vaccine for
CC treatment or prevention of infection by HSV and are represented by the
CC formula P1-x-P2 or P2-x-P1; where P1 = a HSV specific antigenic peptide
CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein
CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and
CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of
CC an immunoprotein which promotes binding to a class of subclass of T cells
CC and which directs a predominantly TH1 type immune response to the peptide
CC P1; and x = a covalent bond or a cleavable or non-cleavable peptide
CC linking group. The conjugated polypeptides can elicit an immune response
CC for neutralizing HSV and killing HSV infected cells. They can be used for
CC the treatment or prevention of HSV infection. In addition, DNA encoding
CC the polypeptide can be used for genetic immunization. They can also be
CC used to diagnose the presence of infection, active or latent, in an
CC individual by HSV by mixing T cells from the individual with the
CC polypeptide and detecting a reaction between the T cells and the
CC polypeptide.
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 36; DB 20; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 20 SSIEFARL 27

RESULT 12
AAE13454
ID AAE13454 standard; peptide; 30 AA.
XX
AC AAE13454;
XX
DT 12-FEB-2002 (first entry)
XX
DE Herpes simplex virus MHC class I peptide antigen #6.
XX
KW Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW major histocompatability complex; MHC; therapy; immune response;
KW malignancy.
XX

OS Herpes simplex virus.
XX
FH Key Location/Qualifiers
FT Region 1..8
FT /note= "Javelin sequence"
FT Region 9..11
FT /note= "Linker"
FT Region 12..19
FT /note= "MHC class I epitope"
FT Region 20..22
FT /note= "Linker"
FT Region 23..30
FT /note= "Javelin sequence"
XX
PN WO200179259-A1.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-US12567.
XX
PR 17-APR-2000; 2000US-197462P.
XX
PA (ROTH/) ROTHMAN J E.
PA (MAYH/) MAYHEW M.
PA (HOEM/) HOE M.
XX
PI Rothman JE, Mayhew M, Hoe M;
XX
DR WPI; 2002-017594/02.
XX
PT A new antigenic complex comprising epitopes non-covalently joined to a
PT heat shock protein by a molecular tether designated a javelin are
PT useful to treat or prevent infectious disease or malignancy -
XX
PS Disclosure; Page 14; 47pp; English.
XX
CC The present invention relates to an antigenic complex, comprising a
CC number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC a tethering molecule referred to as javelin which has affinity for the
CC HSP under physiological conditions, where the epitopes are covalently
CC joined to the tethering molecule and one epitope is major
CC histocompatability complex class I (MHC) and the other MHC class II. The
CC antigenic complex is used to induce immune responses directed towards the
CC treatment or prevention of infectious diseases and malignancies. The
CC present sequence is Herpes simplex virus MHC class I peptide antigen.
XX
SQ Sequence 30 AA;

Query Match 100.0%; Score 36; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 12 SSIEFARL 19

RESULT 13
AAU82093
ID AAU82093 standard; peptide; 38 AA.
XX
AC AAU82093;
XX
DT 09-APR-2002 (first entry)
XX
DE T-cell specific binding ligand peptide construct #27.
XX
KW T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;
KW peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;
KW immunological disorder; immune response; human immunodeficiency virus;
KW herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;
KW acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;
KW autoimmune myocarditis; cytostatic; antiinflammatory.
XX
OS Synthetic.
XX
PN WO200189286-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US16793.
XX
PR 24-MAY-2000; 2000US-206548P.
XX
PA (CELS-) CEL-SCI CORP.
XX
PI Zimmerman DS, Sarin PS;
XX
DR WPI; 2002-083037/11.
XX
PT New T cell binding ligand peptide for treating immunological disorders
PT such as herpes simplex virus, tuberculosis, cancers, acquired
PT immunodeficiency syndrome and allergies -
XX
PS Disclosure; Page 28; 110pp; English.
XX
CC The present invention relates to novel T-cell binding ligand (TCBL)
CC peptides (e.g. peptide G' (modified human MHC class II beta chain
CC peptide G, peptide J (human beta-2-microglobulin peptide) and HIV-1
CC peptides) and TCBL peptide constructs for treating immunological
CC disorders. The peptide constructs are useful for eliciting a cellular
CC immune response in a human patient. The method comprises administering
CC the peptide construct to the patient preferably in combination with an
CC immune response adjuvant. The peptide constructs in the form of
CC conjugated peptides are useful for eliciting a cellular immune response
CC in a patient exposed to or at risk for exposure to the human
CC immunodeficiency virus (HIV). The TCBL peptides are useful for treating a
CC patient suffering from an immunological disorder such as herpes simplex
CC virus (HSV) infection, malaria, tuberculosis, cancers, acquired
CC immunodeficiency syndrome (AIDS), allergies, autoimmune diseases
CC (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune
CC myocarditis, diabetes and lupus) by administering a peptide construct
CC comprising a TCBL peptide bonded to an antigenic peptide associated with
CC the disorder. Unlike prior art peptide conjugates, a modified version of
CC peptide G has long range stabilisation and also enhances the immune
CC response. AAU82019-AAU82114 represent T-cell specific binding ligand
CC peptides, peptide constructs or peptides used in their construction.
XX
SQ Sequence 38 AA;

Query Match 100.0%; Score 36; DB 23; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 23 SSIEFARL 30

RESULT 14
AAU82094
ID AAU82094 standard; peptide; 40 AA.
XX
AC AAU82094;
XX
DT 09-APR-2002 (first entry)
XX
DE T-cell specific binding ligand peptide construct #28.
XX
KW T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;
KW peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;
KW immunological disorder; immune response; human immunodeficiency virus;
KW herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;
KW acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;
KW autoimmune myocarditis; cytostatic; antiinflammatory.
XX

OS Synthetic.
XX
PN WO200189286-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US16793.
XX
PR 24-MAY-2000; 2000US-206548P.
XX
PA (CELS-) CEL-SCI CORP.
XX
PI Zimmerman DS, Sarin PS;
XX
DR WPI; 2002-083037/11.
XX
PT New T cell binding ligand peptide for treating immunological disorders
PT such as herpes simplex virus, tuberculosis, cancers, acquired
PT immunodeficiency syndrome and allergies -
XX
PS Disclosure; Page 28; 110pp; English.
XX
CC The present invention relates to novel T-cell binding ligand (TCBL)
CC peptides (e.g. peptide G' (modified human MHC class II beta chain
CC peptide G, peptide J (human beta-2-microglobulin peptide) and HIV-1
CC peptides) and TCBL peptide constructs for treating immunological
CC disorders. The peptide constructs are useful for eliciting a cellular
CC immune response in a human patient. The method comprises administering
CC the peptide construct to the patient preferably in combination with an
CC immune response adjuvant. The peptide constructs in the form of
CC conjugated peptides are useful for eliciting a cellular immune response
CC in a patient exposed to or at risk for exposure to the human
CC immunodeficiency virus (HIV). The TCBL peptides are useful for treating a
CC patient suffering from an immunological disorder such as herpes simplex
CC virus (HSV) infection, malaria, tuberculosis, cancers, acquired
CC immunodeficiency syndrome (AIDS), allergies, autoimmune diseases
CC (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune
CC myocarditis, diabetes and lupus) by administering a peptide construct
CC comprising a TCBL peptide bonded to an antigenic peptide associated with
CC the disorder. Unlike prior art peptide conjugates, a modified version of
CC peptide G has long range stabilisation and also enhances the immune
CC response. AAU82019-AAU82114 represent T-cell specific binding ligand
CC peptides, peptide constructs or peptides used in their construction.
XX
SQ Sequence 40 AA;

Query Match 100.0%; Score 36; DB 23; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 25 SSIEFARL 32

RESULT 15
AAR14666
ID AAR14666 standard; Protein; 694 AA.
XX
AC AAR14666;
XX
DT 23-JAN-1992 (first entry)
XX
DE Truncated HSVgB polypeptide.
XX
KW Vaccine; antigen.
XX
OS Herpes simplex virus.
XX
PN JP03218397-A.
XX
PD 25-SEP-1991.
XX
PF 21-JUN-1990; 90JP-0161448.
XX
PR 30-NOV-1989; 89JP-0308941.
PR 22-JUN-1989; 89JP-0158238.
PR 21-JUN-1990; 90JP-0161448.
XX
PA (TAKE) TAKEDA CHEMICAL IND KK.
XX
DR WPI; 1991-328397/45.
DR N-PSDB; AAQ14479.
XX
PT HSVgB polypeptide(s) obtd. by recombinant DNA techniques -
PT useful as vaccines against HSV and in diagnosis, can be produced
PT cheaply and safely.
XX
PS Claim 1; Fig 7; 24pp; Japanese.
XX
CC The sequence was deduced from DNA and is that of a truncated HSVgB
CC polypeptide. The recombinant protein can be used to prepare
CC vaccines for prophylaxis of HSV infection and for use in diagnostic
CC kits.
CC See also AAR14665.
XX
SQ Sequence 694 AA;

Query Match 100.0%; Score 36; DB 12; Length 694;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 470 SSIEFARL 477

Search completed: May 22, 2003, 12:08:18
Job time : 33 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:06:12 ; Search time 14 Seconds
(without alignments)
54.934 Million cell updates/sec

Title: US-09-719-494-10
Perfect score: 36
Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	100.0	903	1	VGBEB1	glycoprotein B pre
2	36	100.0	903	1	VGBEK1	glycoprotein B pre
3	36	100.0	904	1	VGBEW7	glycoprotein B pre
4	36	100.0	904	1	VGBEK2	glycoprotein B pre
5	36	100.0	904	1	VGBEB2	glycoprotein B pre
6	35	97.2	885	1	VGBESA	glycoprotein B pre
7	35	97.2	917	1	VGBEBH	glycoprotein B pre
8	31	86.1	920	1	D48349	glycoprotein B pre
9	30	83.3	83	2	T36350	probable regulator
10	30	83.3	153	2	F81429	hypothetical prote
11	30	83.3	344	1	S46020	probable aldehyde
12	30	83.3	350	2	T21106	hypothetical prote
13	30	83.3	465	1	F64630	cysteine-tRNA liga
14	30	83.3	465	2	D71884	cysteine-tRNA liga
15	30	83.3	919	1	VGBEQH	glycoprotein B pre
16	30	83.3	975	2	T42576	probable envelope
17	30	83.3	979	2	JH0109	glycoprotein 14 pr
18	30	83.3	980	1	VGBE2H	glycoprotein B pre
19	30	83.3	980	1	VGBEC6	glycoprotein B pre
20	30	83.3	1315	2	T50262	probable nucleopor
21	29	80.6	154	2	S03242	hypothetical prote
22	29	80.6	208	2	S34238	leukotoxin A - Pas
23	29	80.6	288	1	BVECRA	repA protein - Esc
24	29	80.6	317	2	AF0582	probable permease
25	29	80.6	318	2	C70636	probable echA13 pr
26	29	80.6	321	2	T26974	hypothetical prote
27	29	80.6	452	2	AB3222	two component sens
28	29	80.6	532	2	H84123	CTP synthetase ctr
29	29	80.6	608	2	H81706	oligoendopeptidase
30	29	80.6	784	2	AH2560	hypothetical prote
31	29	80.6	868	1	VGBE31	glycoprotein B - h
32	29	80.6	943	2	B48474	glycoprotein B - f
33	29	80.6	955	1	A35254	leukotoxin A - Pas
34	29	80.6	1344	2	T42637	hypothetical prote
35	28	77.8	208	2	AB2441	hypothetical prote
36	28	77.8	214	2	A70593	probable tmk prote
37	28	77.8	250	1	A41330	acid phosphatase (
38	28	77.8	250	2	AF1025	acid phosphatase (
39	28	77.8	252	2	AE0330	conserved hypothet
40	28	77.8	256	1	WMNVET	EcoRI-T large (ETL
41	28	77.8	277	2	E72564	hypothetical prote
42	28	77.8	291	2	S72848	proteasome beta ch
43	28	77.8	291	2	A70512	probable prcB prot
44	28	77.8	294	2	E70692	hypothetical prote
45	28	77.8	306	2	S27496	nodI protein - Bra

ALIGNMENTS

RESULT 1
VGBEB1
glycoprotein B precursor - human herpesvirus 1 (strain F)
C;Species: human herpesvirus 1
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C;Accession: A03750
R;Pellett, P.E.; Kousoulas, K.G.; Pereira, L.; Roizman, B.
J. Virol. 53, 243-253, 1985
A;Title: Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene: primar
A;Reference number: A03750; MUID:85083254; PMID:2981343
A;Accession: A03750
A;Molecule type: DNA
A;Residues: 1-903 <PEL>
A;Cross-references: GB:M14164; GB:M12398; NID:g330084; PIDN:AAA45776.1; PID:g330086
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-903/Product: glycoprotein B #status predicted <MAT>
F;726-746/Domain: transmembrane #status predicted <TM1>
F;751-771/Domain: transmembrane #status predicted <TM2>
F;774-794/Domain: transmembrane #status predicted <TM3>
F;86,140,254,397,429,477,488,673,818,887/Binding site: carbohydrate (Asn) (covalent
F;115-572,132-528,206-270,363-411,595-632/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
           ||||||||
Db     498 SSIEFARL 505
```

RESULT 2
VGBEK1
glycoprotein B precursor - human herpesvirus 1 (strain KOS)
C;Species: human herpesvirus 1
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C;Accession: A03751
R;Bzik, D.J.; Fox, B.A.; DeLuca, N.A.; Person, S.
Virology 133, 301-314, 1984
A;Title: Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simple
A;Reference number: A03751; MUID:84174058; PMID:6324454
A;Accession: A03751
A;Molecule type: DNA
A;Residues: 1-903 <BZI>
A;Cross-references: GB:K01760; NID:g330082; PIDN:AAA45774.1; PID:g330083
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-903/Product: glycoprotein B #status predicted <MAT>
F;726-746/Domain: transmembrane #status predicted <TM1>

F;751-771/Domain: transmembrane #status predicted <TM2>
F;774-794/Domain: transmembrane #status predicted <TM3>
F;76,86,140,254,397,429,477,488,673,818,887/Binding site: carbohydrate (Asn) (covalent)
F;115-572,132-528,206-270,363-411,595-632/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
           ||||||||
Db     498 SSIEFARL 505
```

RESULT 3
VGBEW7
glycoprotein B precursor - human herpesvirus 1 (strain 17)
C;Species: human herpesvirus 1
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C;Accession: I30084
R;McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perr
J. Gen. Virol. 69, 1531-1574, 1988
A;Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A;Reference number: A30083; MUID:88274327; PMID:2839594
A;Accession: I30084
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-904 <MCG>
A;Cross-references: GB:X14112; GB:D00317; GB:D00374; GB:S40593; NID:g1944536; PIDN:CAA32
C;Genetics:
A;Gene: UL27
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-904/Product: glycoprotein B #status predicted <MAT>
F;727-746/Domain: transmembrane #status predicted <TM1>
F;752-771/Domain: transmembrane #status predicted <TM2>
F;774-795/Domain: transmembrane #status predicted <TM3>
F;87,141,255,398,430,478,489,674,819,888/Binding site: carbohydrate (Asn) (covalent) #st
F;116-573,133-529,207-271,364-412,596-633/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
           ||||||||
Db     499 SSIEFARL 506
```

RESULT 4
VGBEK2
glycoprotein B precursor - human herpesvirus 2 (strain HG52)
C;Species: human herpesvirus 2
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C;Accession: A25611
R;Bzik, D.J.; Debroy, C.; Fox, B.A.; Pederson, N.E.; Person, S.
Virology 155, 322-333, 1986
A;Title: The nucleotide sequence of the gB glycoprotein gene of HSV-2 and comparison wit
A;Reference number: A25611; MUID:87071654; PMID:3024391
A;Accession: A25611
A;Molecule type: DNA
A;Residues: 1-904 <BZI>
A;Cross-references: GB:M14923; NID:g330254; PIDN:AAA66440.1; PID:g330255
R;Norais, N.; Tang, D.; Kaur, S.; Chamberlain, S.H.; Masiarz, F.R.; Burke, R.L.; Marcus,
J. Virol. 70, 7379-7387, 1996
A;Title: Disulfide bonds of herpes simplex virus type 2 glycoprotein gB.
A;Reference number: A58366; MUID:97048015; PMID:8892856
A;Contents: annotation; tryptic peptide disulfide bond assignments
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-904/Product: glycoprotein B #status predicted <MAT>
F;724-744/Domain: transmembrane #status predicted <TM1>
F;749-769/Domain: transmembrane #status predicted <TM2>
F;772-792/Domain: transmembrane #status predicted <TM3>
F;82,136,250,393,425,473,486,671/Binding site: carbohydrate (Asn) (covalent) #status
F;111-570,128-526,202-266,359-407,593-630/Disulfide bonds: #status experimental

Query Match 100.0%; Score 36; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
           ||||||||
Db     496 SSIEFARL 503
```

RESULT 5
VGBEB2
glycoprotein B precursor - human herpesvirus 2 (strain 333)
C;Species: human herpesvirus 2
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C;Accession: A26790; A45928
R;Stuve, L.L.; Brown-Shimer, S.; Pachl, C.; Najarian, R.; Dina, D.; Burke, R.L.
J. Virol. 61, 326-335, 1987
A;Title: Structure and expression of the herpes simplex virus type 2 glycoprotein gB
A;Reference number: A26790; MUID:87112925; PMID:3027364
A;Accession: A26790
A;Molecule type: DNA
A;Residues: 1-904 <STU>
A;Cross-references: GB:M15118; NID:g330256; PIDN:AAA45837.1; PID:g330257
R;Zwaagstra, J.C.; Leung, W.C.
Can. J. Microbiol. 33, 879-887, 1987
A;Title: The nucleotide sequence of herpes simplex virus type 2 (333) glycoprotein gI
A;Reference number: A45928; MUID:88079667; PMID:2446730
A;Accession: A45928
A;Molecule type: DNA
A;Residues: 1-34,'AWPTV',42-307,'T',309-481,'R',483-609,'M',611-664,'R',666-904 <ZWA
A;Cross-references: GB:M24771; NID:g341245; PIDN:AAA60540.1; PID:g623400
A;Note: the authors translated the codon ATG for residue 610 as Ile
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-904/Product: glycoprotein B #status predicted <MAT>
F;724-744/Domain: transmembrane #status predicted <TM1>
F;749-769/Domain: transmembrane #status predicted <TM2>
F;772-792/Domain: transmembrane #status predicted <TM3>
F;82,136,250,393,425,473,486,671/Binding site: carbohydrate (Asn) (covalent) #status
F;111-570,128-526,202-266,359-407,593-630/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
           ||||||||
Db     496 SSIEFARL 503
```

RESULT 6
VGBESA
glycoprotein B precursor - simian herpesvirus SA8 (strain B264)
C;Species: simian herpesvirus SA8
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C;Accession: JQ1332
R;Borchers, K.; Weigelt, W.; Buhk, H.J.; Ludwig, H.; Mankertz, J.
J. Gen. Virol. 72, 2299-2304, 1991
A;Title: Conserved domains of glycoprotein B (gB) of the monkey virus, simian agent
A;Reference number: JQ1332; MUID:91374035; PMID:1895066
A;Accession: JQ1332
A;Molecule type: DNA
A;Residues: 1-885 <BOR>
A;Cross-references: EMBL:X56935; NID:g60438; PIDN:CAA40256.1; PID:g60439
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein

```
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-885/Product: glycoprotein B #status predicted <MAT>
F;712-732/Domain: transmembrane #status predicted <TM1>
F;737-752/Domain: transmembrane #status predicted <TM2>
F;760-780/Domain: transmembrane #status predicted <TM3>
F;68,122,379,411,659/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;97-558,114-514,188-252,345-393,581-618/Disulfide bonds: #status predicted

  Query Match             97.2%;  Score 35;  DB 1;  Length 885;
  Best Local Similarity   87.5%;  Pred. No. 3.9;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||:||||
Db    484 SSVEFARL 491


RESULT 7
VGBEBH
glycoprotein B precursor - bovine herpesvirus 2 (strain BMV)
C;Species: bovine herpesvirus 2
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C;Accession: C29242
R;Hammerschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
Virology 165, 388-405, 1988
A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus t
A;Reference number: A94381; MUID:88306231; PMID:2841793
A;Accession: C29242
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-917 <HAM>
A;Cross-references: GB:M21628; NID:g330752; PIDN:AAA46053.1; PID:g330753
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-917/Product: glycoprotein B #status predicted <GPB>
F;578-594/Domain: transmembrane #status predicted <TM1>
F;770-786/Domain: transmembrane #status predicted <TM2>
F;795-811/Domain: transmembrane #status predicted <TM3>
F;48,110,164,278,421,453,505,564,692/Binding site: carbohydrate (Asn) (covalent) #status

  Query Match             97.2%;  Score 35;  DB 1;  Length 917;
  Best Local Similarity   87.5%;  Pred. No. 4;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||:||||
Db    517 SSVEFARL 524


RESULT 8
D48349
glycoprotein B precursor - saimiriine herpesvirus 1 (strain MV-5-4-PSL)
C;Species: saimiriine herpesvirus 1
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Sep-1994
C;Accession: D48349
R;Eberle, R.; Black, D.
Arch. Virol. 129, 167-182, 1993
A;Title: Sequence analysis of herpes simplex virus gB gene homologs of two platyrrhine m
A;Reference number: A48349; MUID:93228440; PMID:8385913
A;Accession: D48349
A;Molecule type: DNA
A;Residues: 1-920 <EBE>
A;Note: sequence extracted from NCBI backbone (NCBIN:129066, NCBIP:129068)
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-920/Product: glycoprotein B #status predicted <MAT>
F;576-592/Region: hydrophobic
F;761-780/Domain: transmembrane #status predicted <TM1>
F;787-804/Domain: transmembrane #status predicted <TM2>
F;98,119,152,409,441,683,733/Binding site: carbohydrate (Asn) (covalent) #status predict

  Query Match             86.1%;  Score 31;  DB 1;  Length 920;
  Best Local Similarity   85.7%;  Pred. No. 39;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 SIEFARL 8
          |:|||||
Db    509 SVEFARL 515


RESULT 9
T36350
probable regulator - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36350
R;Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
submitted to the EMBL Data Library, May 1999
A;Reference number: Z21575
A;Accession: T36350
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-83 <OLI>
A;Cross-references: EMBL:AL049841; PIDN:CAB42777.1; GSPDB:GN00070; SCOEDB:SCE9.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE9.31c

  Query Match             83.3%;  Score 30;  DB 2;  Length 83;
  Best Local Similarity   75.0%;  Pred. No. 4.7;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          | :|||||
Db     33 SCVEFARL 40


RESULT 10
F81429
hypothetical protein Cj0126c [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: F81429
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Ch
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; B
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveal
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: F81429
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <PAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72610.1; PID:g6
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0126c
C;Superfamily: conserved hypothetical protein HI0033

  Query Match             83.3%;  Score 30;  DB 2;  Length 153;
  Best Local Similarity   75.0%;  Pred. No. 9.3;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          :|||||:|
Db     83 TSIEFAKL 90


RESULT 11
S46020
probable aldehyde reductase (EC 1.1.1.21) - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YBR1127; hypothetical protein YBR149w
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002
C;Accession: S46020; S28668
```

```
R;Entian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.; Li, Z.; Niegemann, E.; Sche
 S.; Gruenbein, R.; Hedges, D.; Kiesau, P.; Korol, S.; Krems, B.; Proft, M.; Siegers, K.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S46013
A;Accession: S46020
A;Molecule type: DNA
A;Residues: 1-344 <ENT>
A;Cross-references: EMBL:Z36018; NID:g536473; PIDN:CAA85107.1; PID:g536474; GSPDB:GN0000
A;Experimental source: strain S288C
R;Martinez-Soriano, J.P.; Wong, W.M.; van Ryk, D.I.; Nazar, R.N.
J. Mol. Biol. 217, 629-635, 1991
A;Title: A widely distributed "cat" family of repetitive DNA sequences.
A;Reference number: S28668; MUID:91171289; PMID:2005616
A;Accession: S28668
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-55,'LQSKLDN',63-70,'SR',74-344 <MAR>
A;Cross-references: EMBL:M95580; NID:g172583; PIDN:AAA35037.1; PID:g172584
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C;Genetics:
A;Gene: SGD:ARA1; MIPS:YBR149w
A;Cross-references: SGD:S0000353
A;Map position: 2R
C;Superfamily: aldehyde reductase
C;Keywords: oxidoreductase

  Query Match             83.3%;  Score 30;  DB 1;  Length 344;
  Best Local Similarity   87.5%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          |||||| |
Db    294 SSIEFASL 301


RESULT 12
T21106
hypothetical protein F19B6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21106
R;Thomas, K.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19375
A;Accession: T21106
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-350 <WIL>
A;Cross-references: EMBL:Z69635; NID:g1200023; PIDN:CAA93458.1; GSPDB:GN00022; CESP:F19E
A;Experimental source: clone F19B6
C;Genetics:
A;Gene: CESP:F19B6.3
A;Map position: 4
A;Introns: 180/3; 274/2

  Query Match             83.3%;  Score 30;  DB 2;  Length 350;
  Best Local Similarity   62.5%;  Pred. No. 23;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          |::|||:
Db    125 SAVEFARI 132


RESULT 13
F64630
cysteine-tRNA ligase (EC 6.1.1.16) - Helicobacter pylori (strain 26695)
N;Alternate names: cysteinyl-tRNA synthetase
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 03-Jun-2002
C;Accession: F64630; C53739
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64630
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-465 <TOM>
A;Cross-references: GB:AE000511; NID:g2314019; PIDN:AAD07934.1; PID:g2314022; TIGR:HI
A;Experimental source: strain 26695
R;Cover, T.L.; Tummuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.
J. Biol. Chem. 269, 10566-10573, 1994
A;Title: Divergence of genetic sequences for the vacuolating cytotoxin among Helicob
A;Reference number: A53739; MUID:94193753; PMID:8144644
A;Accession: C53739
A;Molecule type: DNA
A;Residues: 278-279,'I',281-331,'N',333-391,'V',393-431,'Q',433-439,'H',441-465 <COV
A;Cross-references: GB:U05676; NID:g471727; PIDN:AAA17656.1; PID:g471728
A;Experimental source: strain ATCC 49503
C;Superfamily: cysteine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

  Query Match             83.3%;  Score 30;  DB 1;  Length 465;
  Best Local Similarity   75.0%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||| |:
Db    159 SSIEFGRI 166


RESULT 14
D71884
cysteine-tRNA ligase (EC 6.1.1.16) - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C;Accession: D71884
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: D71884
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-465 <ARN>
A;Cross-references: GB:AE001511; GB:AE001439; NID:g4155382; PIDN:AAD06399.1; PID:g41
A;Experimental source: strain J99
C;Genetics:
A;Gene: cysS
C;Superfamily: cysteine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

  Query Match             83.3%;  Score 30;  DB 2;  Length 465;
  Best Local Similarity   62.5%;  Pred. No. 32;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||:||:|:
Db    159 SSVEFSRI 166


RESULT 15
VGBEQH
glycoprotein B precursor - equine herpesvirus 4 (strain 1942)
C;Species: equine herpesvirus 4
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C;Accession: A31880
R;Riggio, M.P.; Cullinane, A.A.; Onions, D.E.
J. Virol. 63, 1123-1133, 1989
A;Title: Identification and nucleotide sequence of the glycoprotein gB gene of equine
```

A;Reference number: A31880; MUID:89125704; PMID:2915378
A;Accession: A31880
A;Molecule type: DNA
A;Residues: 1-919 <RIG>
A;Cross-references: GB:M26171; NID:g341446; PIDN:AAA46106.1; PID:g514920
C;Superfamily: herpesvirus glycoprotein B
C;Keywords: glycoprotein; transmembrane protein
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-919/Product: glycoprotein B #status predicted <GPB>
F;740-809/Domain: transmembrane #status predicted <TMN>
F;106,216,321,364,438,456,493,499,666,688/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 83.3%; Score 30; DB 1; Length 919;
Best Local Similarity 87.5%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 SSIEFARL 8
             |||||| |
Db     512 SSIEFAML 519
```

Search completed: May 22, 2003, 12:10:11
Job time : 16 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:57 ; Search time 7.5 Seconds
(without alignments)
44.241 Million cell updates/sec

Title: US-09-719-494-10
Perfect score: 36
Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	100.0	903	1	VGLB_HSV1F	P06436 herpes simp
2	36	100.0	904	1	VGLB_HSV11	P10211 herpes simp
3	36	100.0	904	1	VGLB_HSV1K	P06437 herpes simp
4	36	100.0	904	1	VGLB_HSV1P	P08665 herpes simp
5	36	100.0	904	1	VGLB_HSV23	P06763 herpes simp
6	36	100.0	904	1	VGLB_HSV2H	P08666 herpes simp
7	35	97.2	885	1	VGLB_HSV2S	P24994 herpes simp
8	35	97.2	917	1	VGLB_HSVB2	P12641 bovine herp
9	31	86.1	920	1	VGLB_HSVSM	Q04464 herpesvirus
10	30	83.3	344	1	ARA1_YEAST	P38115 saccharomyc
11	30	83.3	435	1	ERF1_TETTH	Q9u8u5 tetrahymena
12	30	83.3	445	1	ERF1_OXYTR	Q9bmx3 oxytricha t
13	30	83.3	445	1	ERF1_STYLE	Q9bmm0 stylonychia
14	30	83.3	445	1	ERF1_STYMT	Q9bmm1 stylonychia
15	30	83.3	457	1	ERF1_GIALA	Q9ncp1 giardia lam
16	30	83.3	465	1	SYC_HELPJ	Q9zkw6 helicobacte
17	30	83.3	465	1	SYC_HELPY	P41259 helicobacte
18	30	83.3	919	1	VGLB_HSVE4	P17472 equine herp
19	30	83.3	979	1	VGLB_HSVEL	P25218 equine herp
20	30	83.3	980	1	VGLB_HSVE1	P18050 equine herp
21	30	83.3	980	1	VGLB_HSVEA	P18551 equine herp
22	30	83.3	980	1	VGLB_HSVEB	P28922 equine herp
23	29	80.6	154	1	Y17K_SSV1	P20196 sulfolobus
24	29	80.6	288	1	REPA_ECOLI	P05833 escherichia
25	29	80.6	317	1	KDT2_SALTI	Q8z8h3 salmonella
26	29	80.6	868	1	VGLB_VZVD	P09257 varicella-z
27	29	80.6	953	1	LKA3_PASHA	P55116 pasteurella
28	29	80.6	955	1	LKAA_PASHA	P55117 pasteurella
29	29	80.6	1344	1	IF3A_MOUSE	P23116 mus musculu
30	29	80.6	1382	1	IF3A_HUMAN	Q14152 homo sapien
31	28	77.8	250	1	PHON_SALTI	Q934j6 salmonella
32	28	77.8	250	1	PHON_SALTY	P26976 salmonella
33	28	77.8	256	1	PCNA_NPVAC	P11038 autographa
34	28	77.8	304	1	NODI_RHIS3	P72335 rhizobium s
35	28	77.8	306	1	NODI_BRAJA	P26050 bradyrhizob
36	28	77.8	311	1	NODI_RHILV	P08720 rhizobium l
37	28	77.8	312	1	VPB_BPMU	P03763 bacteriopha
38	28	77.8	340	1	NODI_RHILO	P23703 rhizobium l
39	28	77.8	355	1	NODI_RHIME	O52618 rhizobium m
40	28	77.8	538	1	RBB5_HUMAN	Q15291 homo sapien
41	28	77.8	552	1	SYQ_CLOPE	Q8xmp3 clostridium
42	28	77.8	660	1	ALIA_STRPN	P35592 streptococc
43	28	77.8	750	1	CBBB_BACTV	Q9ziu5 bacillus th
44	28	77.8	933	1	VGLB_HSVA1	Q04463 herpesvirus
45	28	77.8	4466	1	DYHC_ANTCR	P39057 anthocidari

ALIGNMENTS

RESULT 1
VGLB_HSV1F
ID VGLB_HSV1F STANDARD; PRT; 903 AA.
AC P06436;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein B precursor.
GN GB OR UL27.
OS Herpes simplex virus (type 1 / strain F).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10304;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85083254; PubMed=2981343;
RA Pellett P.E., Kousoulas K.G., Pereira L., Roizman B.;
RT "Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene:
RT primary sequence and predicted protein structure of the wild type and
RT of monoclonal antibody-resistant mutants.";
RL J. Virol. 53:243-253(1985).
RN [2]
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hammerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA Ludwig H.;
RT "Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).";
RL Virology 165:406-418(1988).
CC -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC GB, GC, GG, GD, GI, AND GE.
CC -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC REQUIRED FOR VIRAL GROWTH.
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC ----------

CC ----------
DR EMBL; M14164; AAA45776.1; -.
DR EMBL; M21633; AAA45788.1; -.
DR PIR; A03750; VGBEB1.
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B; 1.
DR ProDom; PD000693; Glycoprot_B; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 29
FT CHAIN 30 903 GLYCOPROTEIN B.
FT DOMAIN 31 729 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 730 745 POTENTIAL.

```
FT   TRANSMEM    751    770       POTENTIAL.
FT   TRANSMEM    774    794       POTENTIAL.
FT   DOMAIN      795    903       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     86     86       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    397    397       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    429    429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    488    488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    673    673       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   903 AA;  100104 MW;  73BDCA7813DB35E8 CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 903;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     498 SSIEFARL 505


RESULT 2
VGLB_HSV11
ID   VGLB_HSV11     STANDARD;      PRT;   904 AA.
AC   P10211;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC       GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10879; BAA01673.1; -.
DR   EMBL; X14112; CAA32320.1; -.
DR   PIR; I30084; VGBEW7.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     30
FT   CHAIN        31    904       GLYCOPROTEIN B.
FT   DOMAIN       31    730       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    731    746       POTENTIAL.
FT   TRANSMEM    752    771       POTENTIAL.
FT   TRANSMEM    775    795       POTENTIAL.
FT   DOMAIN      796    904       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    141    141       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    674    674       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   904 AA;  100292 MW;  2C14E8B1284C1F3A CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     499 SSIEFARL 506


RESULT 3
VGLB_HSV1K
ID   VGLB_HSV1K     STANDARD;      PRT;   904 AA.
AC   P06437;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27.
OS   Herpes simplex virus (type 1 / strain KOS).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10306;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84174058; PubMed=6324454;
RA   Bzik D.J., Fox B.A., Deluca N.A., Person S.;
RT   "Nucleotide sequence specifying the glycoprotein gene, gB, of herpes
RT   simplex virus type 1.";
RL   Virology 133:301-314(1984).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE=87071654; PubMed=3024391;
RA   Bzik D.J., Debroy C., Fox B.A., Pederson N.E., Person S.;
RT   "The nucleotide sequence of the gB glycoprotein gene of HSV-2 and
RT   comparison with the corresponding gene of HSV-1.";
RL   Virology 155:322-333(1986).
RN   [3]
RP   REVISIONS.
RA   Pederson N.E.;
RL   Submitted (APR-1987) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC       GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K01760; AAA45774.1; -.
DR   PIR; A03751; VGBEK1.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     30
FT   CHAIN        31    904       GLYCOPROTEIN B.
FT   DOMAIN       31    730       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    731    746       POTENTIAL.
FT   TRANSMEM    752    771       POTENTIAL.
FT   TRANSMEM    775    795       POTENTIAL.
FT   DOMAIN      796    904       CYTOPLASMIC (POTENTIAL).
```

```
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    141    141       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    674    674       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   904 AA;  100368 MW;  B97B7F8DE5FBA299 CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 4
VGLB_HSV1P
ID   VGLB_HSV1P     STANDARD;      PRT;   904 AA.
AC   P08665;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27.
OS   Herpes simplex virus (type 1 / strain Patton).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10308;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87112925; PubMed=3027364;
RA   Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RA   Burke R.L.;
RT   "Structure and expression of the herpes simplex virus type 2
RT   glycoprotein gB gene.";
RL   J. Virol. 61:326-335(1987).
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC       GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03541; AAA45778.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL       1     30
FT   CHAIN       31    904       GLYCOPROTEIN B.
FT   DOMAIN      31    730       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   731    746       POTENTIAL.
FT   TRANSMEM   752    772       POTENTIAL.
FT   TRANSMEM   775    795       POTENTIAL.
FT   DOMAIN     796    904       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   141    141       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   674    674       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   904 AA;  100115 MW;  7825E1DC830A626F CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 5
VGLB_HSV23
ID   VGLB_HSV23     STANDARD;      PRT;   904 AA.
AC   P06763;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27 OR GB2.
OS   Herpes simplex virus (type 2 / strain 333).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10313;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87112925; PubMed=3027364;
RA   Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RA   Burke R.L.;
RT   "Structure and expression of the herpes simplex virus type 2
RT   glycoprotein gB gene.";
RL   J. Virol. 61:326-335(1987).
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC       2: GH, GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15118; AAA45837.1; -.
DR   PIR; A26790; VGBEB2.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    904       GLYCOPROTEIN B.
FT   DOMAIN      23    727       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   728    743       POTENTIAL.
FT   TRANSMEM   749    768       POTENTIAL.
FT   TRANSMEM   772    792       POTENTIAL.
FT   DOMAIN     793    904       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   393    393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   486    486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   671    671       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   904 AA;  100186 MW;  A8B36F74FDBCB539 CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    496 SSIEFARL 503
```

```
RESULT 6
VGLB_HSV2H
ID   VGLB_HSV2H     STANDARD;      PRT;   904 AA.
AC   P08666; P89450;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27.
OS   Herpes simplex virus (type 2 / strain HG52).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10315;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87071654; PubMed=3024391;
RA   Bzik D.J., Debroy C., Fox B.A., Pederson N.E., Person S.;
RT   "The nucleotide sequence of the gB glycoprotein gene of HSV-2 and
RT   comparison with the corresponding gene of HSV-1.";
RL   Virology 155:322-333(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dolan A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC       2: GH, GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14923; AAA66440.1; -.
DR   EMBL; Z86099; CAB06752.1; -.
DR   PIR; A25611; VGBEK2.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    904       GLYCOPROTEIN B.
FT   DOMAIN       23    727       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    728    743       POTENTIAL.
FT   TRANSMEM    749    768       POTENTIAL.
FT   TRANSMEM    772    792       POTENTIAL.
FT   DOMAIN      793    904       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    393    393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    486    486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    671    671       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     92     92       L -> V (IN REF. 1).
FT   CONFLICT    198    198       T -> A (IN REF. 1).
FT   CONFLICT    308    308       S -> T (IN REF. 1).
FT   CONFLICT    438    438       L -> Q (IN REF. 1).
FT   CONFLICT    568    568       S -> A (IN REF. 1).
FT   CONFLICT    619    620       EL -> DV (IN REF. 1).
FT   CONFLICT    636    636       R -> G (IN REF. 1).
SQ   SEQUENCE   904 AA;  100217 MW;  AB050A3AFB4F1066 CRC64;

  Query Match             100.0%;  Score 36;  DB 1;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     496 SSIEFARL 503


RESULT 7
VGLB_HSV2S
ID   VGLB_HSV2S     STANDARD;      PRT;   885 AA.
AC   P24994;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB OR UL27.
OS   Herpes simplex virus (type 2 / strain SA8) (Simian agent 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B264;
RX   MEDLINE=91374035; PubMed=1895066;
RA   Borchers K., Weigelt W., Buhk H.-J., Ludwig H., Mankertz J.;
RT   "Conserved domains of glycoprotein B (gB) of the monkey virus, simian
RT   agent 8, identified by comparison with herpesvirus gBs.";
RL   J. Gen. Virol. 72:2299-2304(1991).
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC       2: GH, GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56935; CAA40256.1; -.
DR   PIR; JQ1332; VGBESA.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    885       GLYCOPROTEIN B.
FT   DOMAIN       35    715       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    716    731       POTENTIAL.
FT   TRANSMEM    737    756       POTENTIAL.
FT   TRANSMEM    760    780       POTENTIAL.
FT   DOMAIN      781    885       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    379    379       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   885 AA;  97811 MW;  39E495B329AB94E4 CRC64;

  Query Match             97.2%;  Score 35;  DB 1;  Length 885;
  Best Local Similarity   87.5%;  Pred. No. 1.7;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||:||||
Db     484 SSVEFARL 491


RESULT 8
VGLB_HSVB2
```

```
ID   VGLB_HSVB2     STANDARD;      PRT;   917 AA.
AC   P12641;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B-1 precursor.
OS   Bovine herpesvirus type 2 (strain BMV) (Bovine mammillitis virus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88306231; PubMed=2841793;
RA   Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
RA   Buhk H.-J.;
RT   "Conservation of a gene cluster including glycoprotein B in bovine
RT   herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL   Virology 165:388-405(1988).
RN   [2]
RP   SEQUENCE OF 1-200 FROM N.A.
RX   MEDLINE=88306232; PubMed=2457278;
RA   Hammerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA   Ludwig H.;
RT   "Common epitopes of glycoprotein B map within the major DNA-binding
RT   proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT   virus type 1 (HSV-1).";
RL   Virology 165:406-418(1988).
CC   -!- FUNCTION: GB1 IS A 130 kDa GLYCOPROTEIN WHICH IS NECESSARY FOR THE
CC       PENETRATION OF THE VIRUS INTO THE HOST CELL AND THE INDUCTION OF A
CC       SYNCYTIAL PHENOTYPE.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21628; AAA46053.1; -.
DR   EMBL; M21632; AAA46052.1; -.
DR   PIR; C29242; VGBEBH.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    917       GLYCOPROTEIN B-1.
FT   TRANSMEM    578    594       POTENTIAL.
FT   TRANSMEM    770    786       POTENTIAL.
FT   TRANSMEM    795    811       POTENTIAL.
FT   CARBOHYD     48     48       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    453    453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    505    505       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   917 AA;  101882 MW;  1B96CBF50DB4D3F3 CRC64;

  Query Match             97.2%;  Score 35;  DB 1;  Length 917;
  Best Local Similarity   87.5%;  Pred. No. 1.8;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
            ||:|||||
Db     517 SSVEFARL 524


RESULT 9
```

```
VGLB_HSVSM
ID   VGLB_HSVSM     STANDARD;      PRT;   920 AA.
AC   Q04464;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glycoprotein B precursor.
GN   UL27.
OS   Herpesvirus saimiri (type 1 / strain MV-5-4-PSL) (Marmoset
OS   herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae.
OX   NCBI_TaxID=10353;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93228440; PubMed=8385913;
RA   Eberle R., Black D.;
RT   "Sequence analysis of herpes simplex virus gB gene homologs of two
RT   platyrrhine monkey alpha-herpesviruses.";
RL   Arch. Virol. 129:167-182(1993).
CC   -!- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC       REQUIRED FOR VIRAL GROWTH.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95786; AAA43841.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    920       GLYCOPROTEIN B.
FT   DOMAIN       29    739       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    740    755       POTENTIAL.
FT   TRANSMEM    761    781       POTENTIAL.
FT   TRANSMEM    784    804       POTENTIAL.
FT   DOMAIN      805    920       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    441    441       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    733    733       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   920 AA;  101661 MW;  62C4D0315D8C2DB7 CRC64;

  Query Match             86.1%;  Score 31;  DB 1;  Length 920;
  Best Local Similarity   85.7%;  Pred. No. 17;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 SIEFARL 8
            |:|||||
Db     509 SVEFARL 515


RESULT 10
ARA1_YEAST
ID   ARA1_YEAST     STANDARD;      PRT;   344 AA.
AC   P38115;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   D-arabinose dehydrogenase [NAD(P)+] heavy chain (EC 1.1.1.117).
GN   ARA1 OR YBR149W OR YBR1127.
```

```
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Entian K.-D., Koetter P., Rose M., Becker J., Grey M., Li Z.,
RA   Niegemann E., Schenk-Groeninger R., Servos J., Wehner E.,
RA   Wolter R., Brendel M., Bauer J., Braun H., Dern K., Duesterhus S.,
RA   Gruenbein R., Hedges D., Kiesau P., Korol S., Krems B., Proft M.,
RA   Siegers K., Baur A., Boles E., Miosga T.,
RA   Schaaff-Gerstenschlaeger I., Zimmermann F.K.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91171289; PubMed=2005616;
RA   Martinez-Soriano J.P., Wong W.M., van Ryk D.I., Nazar R.N.;
RT   "A widely distributed 'CAT' family of repetitive DNA sequences.";
RL   J. Mol. Biol. 217:629-635(1991).
RN   [3]
RP   CHARACTERIZATION, AND SEQUENCE OF 7-20.
RX   MEDLINE=99117109; PubMed=9920381;
RA   Kim S.T., Huh W.K., Lee B.H., Kang S.O.;
RT   "D-arabinose dehydrogenase and its gene from Saccharomyces
RT   cerevisiae.";
RL   Biochim. Biophys. Acta 1429:29-39(1998).
CC   -!- FUNCTION: CATALYZES THE OXIDATION OF D-ARABINOSE, L-XYLOSE, L-
CC       FUCOSE AND L-GALACTOSE IN THE PRESENCE OF NADP+.
CC   -!- CATALYTIC ACTIVITY: D-arabinose + NAD(P)(+) = D-arabinono-1,4-
CC       lactone + NAD(P)H.
CC   -!- SUBUNIT: HETERODIMER OF A HEAVY CHAIN AND A LIGHT CHAIN.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- MISCELLANEOUS: EXHIBITS MAXIMUM ACTIVITY AT PH 10.0 AND AROUND 30
CC       DEGREES CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36018; CAA85107.1; -.
DR   EMBL; M95580; AAA35037.1; -.
DR   PIR; S46020; S46020.
DR   HSSP; P06632; 1HW6.
DR   SGD; S0000353; ARA1.
DR   InterPro; IPR001395; Aldo/ket_red.
DR   Pfam; PF00248; aldo_ket_red; 1.
DR   PRINTS; PR00069; ALDKETRDTASE.
DR   ProDom; PD000288; Aldo/ket_red; 1.
DR   PROSITE; PS00798; ALDOKETO_REDUCTASE_1; 1.
DR   PROSITE; PS00062; ALDOKETO_REDUCTASE_2; 1.
DR   PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
KW   Oxidoreductase; NAD.
FT   ACT_SITE     71     71       HYDROGEN-BOND DONOR (BY SIMILARITY).
FT   CONFLICT     56     62       AAIKAGY -> LQSKLDN (IN REF. 2).
FT   CONFLICT     71     73       YET -> SR (IN REF. 2).
SQ   SEQUENCE   344 AA;  38883 MW;  F9B6D9333B18FEEC CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 344;
  Best Local Similarity   87.5%;  Pred. No. 9.8;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           |||||| |
Db     294 SSIEFASL 301


RESULT 11
ERF1_TETTH
ID   ERF1_TETTH     STANDARD;      PRT;   435 AA.
AC   Q9U8U5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic peptide chain release factor subunit 1 (eRF1) (Eukaryotic
DE   release factor 1).
GN   ERF1.
OS   Tetrahymena thermophila.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99402978; PubMed=10471834;
RA   Karamyshev A.L., Ito K., Nakamura Y.;
RT   "Polypeptide release factor eRF1 from Tetrahymena thermophila: cDNA
RT   cloning, purification and complex formation with yeast eRF3.";
RL   FEBS Lett. 457:483-488(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21113188; PubMed=11160924;
RA   Inagaki Y., Doolittle W.F.;
RT   "Class I release factors in ciliates variant genetic codes.";
RL   Nucleic Acids Res. 29:921-927(2001).
CC   -!- FUNCTION: Directs the termination of nascent peptide synthesis
CC       (translation) in response to the termination codon UGA.
CC       In T.thermophila UAA and UAG codes for glutamine.
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB026195; BAA85336.1; -.
DR   EMBL; AF298833; AAK07831.1; -.
DR   HSSP; P46055; 1DT9.
DR   InterPro; IPR004403; eRF1.
DR   InterPro; IPR005140; eRF1_1.
DR   InterPro; IPR005141; eRF1_2.
DR   InterPro; IPR005142; eRF1_3.
DR   Pfam; PF03463; eRF1_1; 1.
DR   Pfam; PF03464; eRF1_2; 1.
DR   Pfam; PF03465; eRF1_3; 1.
DR   TIGRFAMs; TIGR00108; eRF; 1.
KW   Protein biosynthesis.
SQ   SEQUENCE   435 AA;  49558 MW;  89013ED4C8612646 CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 435;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||: ||||
Db     185 SSVRFARL 192


RESULT 12
ERF1_OXYTR
ID   ERF1_OXYTR     STANDARD;      PRT;   445 AA.
AC   Q9BMX3; Q9BMM2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic peptide chain release factor subunit 1 (eRF1) (Eukaryotic
DE   release factor 1).
```

```
GN   ERF1.
OS   Oxytricha trifallax.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
OC   Stichotrichida; Oxytrichidae; Oxytricha.
OX   NCBI_TaxID=5946;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21113188; PubMed=11160924;
RA   Inagaki Y., Doolittle W.F.;
RT   "Class I release factors in ciliates variant genetic codes.";
RL   Nucleic Acids Res. 29:921-927(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21154807; PubMed=11231122;
RA   Lozupone C.A., Knight R.D., Landweber L.F.;
RT   "The molecular basis of nuclear genetic code change in ciliates.";
RL   Curr. Biol. 11:65-74(2001).
CC   -!- FUNCTION: Directs the termination of nascent peptide synthesis
CC       (translation) in response to the termination codon UGA.
CC       In O.trifallax UAA and UAG codes for glutamine.
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF298830; AAK07828.1; -.
DR   EMBL; AF317832; AAK12090.1; -.
DR   HSSP; P46055; 1DT9.
DR   InterPro; IPR004403; eRF1.
DR   InterPro; IPR005140; eRF1_1.
DR   InterPro; IPR005141; eRF1_2.
DR   InterPro; IPR005142; eRF1_3.
DR   Pfam; PF03463; eRF1_1; 1.
DR   Pfam; PF03464; eRF1_2; 1.
DR   Pfam; PF03465; eRF1_3; 1.
DR   TIGRFAMs; TIGR00108; eRF; 1.
KW   Protein biosynthesis.
SQ   SEQUENCE   445 AA;  49716 MW;  7D637F4C2572A09F CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 445;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||: ||||
Db     191 SSVRFARL 198


RESULT 13
ERF1_STYLE
ID   ERF1_STYLE     STANDARD;      PRT;   445 AA.
AC   Q9BMM0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic peptide chain release factor subunit 1 (eRF1) (Eukaryotic
DE   release factor 1).
GN   ERF1.
OS   Stylonychia lemnae.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
OC   Stichotrichida; Oxytrichidae; Stylonychia.
OX   NCBI_TaxID=5949;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21154807; PubMed=11231122;
RA   Lozupone C.A., Knight R.D., Landweber L.F.;
RT   "The molecular basis of nuclear genetic code change in ciliates.";
RL   Curr. Biol. 11:65-74(2001).
CC   -!- FUNCTION: Directs the termination of nascent peptide synthesis
CC       (translation) in response to the termination codon UGA.
CC       In Stylonchia UAA and UAG codes for glutamine.
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF317834; AAK12092.1; -.
DR   HSSP; P46055; 1DT9.
DR   InterPro; IPR004403; eRF1.
DR   InterPro; IPR005140; eRF1_1.
DR   InterPro; IPR005141; eRF1_2.
DR   InterPro; IPR005142; eRF1_3.
DR   Pfam; PF03463; eRF1_1; 1.
DR   Pfam; PF03464; eRF1_2; 1.
DR   Pfam; PF03465; eRF1_3; 1.
DR   TIGRFAMs; TIGR00108; eRF; 1.
KW   Protein biosynthesis.
SQ   SEQUENCE   445 AA;  49558 MW;  CCE0D191E15D74EC CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 445;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||: ||||
Db     191 SSVRFARL 198


RESULT 14
ERF1_STYMT
ID   ERF1_STYMT     STANDARD;      PRT;   445 AA.
AC   Q9BMM1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic peptide chain release factor subunit 1 (eRF1) (Eukaryotic
DE   release factor 1).
GN   ERF1.
OS   Stylonychia mytilus.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
OC   Stichotrichida; Oxytrichidae; Stylonychia.
OX   NCBI_TaxID=5950;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21154807; PubMed=11231122;
RA   Lozupone C.A., Knight R.D., Landweber L.F.;
RT   "The molecular basis of nuclear genetic code change in ciliates.";
RL   Curr. Biol. 11:65-74(2001).
CC   -!- FUNCTION: Directs the termination of nascent peptide synthesis
CC       (translation) in response to the termination codon UGA.
CC       In Stylonchia UAA and UAG codes for glutamine.
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AF317833; AAK12091.1; -.
DR   HSSP; P46055; 1DT9.
DR   InterPro; IPR004403; eRF1.
DR   InterPro; IPR005140; eRF1_1.
DR   InterPro; IPR005141; eRF1_2.
DR   InterPro; IPR005142; eRF1_3.
DR   Pfam; PF03463; eRF1_1; 1.
DR   Pfam; PF03464; eRF1_2; 1.
DR   Pfam; PF03465; eRF1_3; 1.
DR   TIGRFAMs; TIGR00108; eRF; 1.
KW   Protein biosynthesis.
SQ   SEQUENCE   445 AA;  49670 MW;  0F3622FEA1BD50B8 CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 445;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||: ||||
Db    191 SSVRFARL 198


RESULT 15
ERF1_GIALA
ID   ERF1_GIALA     STANDARD;      PRT;   457 AA.
AC   Q9NCP1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic peptide chain release factor subunit 1 (eRF1) (Eukaryotic
DE   release factor 1).
GN   ERF1.
OS   Giardia lamblia (Giardia intestinalis).
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX   NCBI_TaxID=5741;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20293306; PubMed=10833194;
RA   Inagaki Y., Doolittle W.F.;
RT   "Evolution of the eukaryotic translation termination system: origins
RT   of release factors.";
RL   Mol. Biol. Evol. 17:882-889(2000).
CC   -!- FUNCTION: Directs the termination of nascent peptide synthesis
CC       (translation) in response to the termination codons UAA, UAG and
CC       UGA (By similarity).
CC   -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF198107; AAF74402.1; -.
DR   HSSP; P46055; 1DT9.
DR   InterPro; IPR004403; eRF1.
DR   InterPro; IPR005140; eRF1_1.
DR   InterPro; IPR005141; eRF1_2.
DR   InterPro; IPR005142; eRF1_3.
DR   Pfam; PF03463; eRF1_1; 1.
DR   Pfam; PF03464; eRF1_2; 1.
DR   Pfam; PF03465; eRF1_3; 1.
DR   TIGRFAMs; TIGR00108; eRF; 1.
KW   Protein biosynthesis.
SQ   SEQUENCE   457 AA;  51029 MW;  EE185FFFD1F0C943 CRC64;

  Query Match             83.3%;  Score 30;  DB 1;  Length 457;
  Best Local Similarity   75.0%;  Pred. No. 14;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||: ||||
Db    191 SSVRFARL 198


Search completed: May 22, 2003, 12:08:39
Job time : 8.5 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:04:33 ; Search time 25 Seconds
(without alignments)
65.935 Million cell updates/sec

Title: US-09-719-494-10
Perfect score: 36
Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	100.0	583	12	Q68981	Q68981 human herpe
2	36	100.0	887	12	Q9YW93	Q9yw93 macropodid
3	36	100.0	893	12	Q65540	Q65540 baboon herp
4	36	100.0	894	12	Q9YW92	Q9yw92 macropodid
5	36	100.0	901	12	Q69464	Q69464 herpes simp
6	36	100.0	901	12	Q69465	Q69465 herpes simp
7	36	100.0	902	12	Q69095	Q69095 herpes simp
8	36	100.0	903	12	Q69076	Q69076 human herpe
9	36	100.0	904	12	Q69526	Q69526 human herpe
10	36	100.0	904	12	Q9QLM8	Q9qlm8 human herpe
11	36	100.0	904	12	Q9IWU4	Q9iwu4 human herpe
12	36	100.0	904	12	Q9DXE4	Q9dxe4 herpes simp
13	36	100.0	904	12	O37453	O37453 herpes simp
14	36	100.0	904	12	Q91C63	Q91c63 herpes simp
15	36	100.0	904	12	Q89920	Q89920 herpes simp
16	35	97.2	885	12	Q69387	Q69387 cercopithec
17	35	97.2	891	12	Q66018	Q66018 simian herp
18	35	97.2	891	12	Q9J052	Q9j052 simian herp
19	35	97.2	908	12	Q65538	Q65538 baboon herp
20	31	86.1	261	17	Q8TJH6	Q8tjh6 methanosarc
21	31	86.1	429	5	Q8WQT3	Q8wqt3 trichomonas
22	30	83.3	83	16	Q9X8K3	Q9x8k3 streptomyce
23	30	83.3	153	16	Q9PJ01	Q9pj01 campylobact
24	30	83.3	350	5	Q19582	Q19582 caenorhabdi
25	30	83.3	386	5	Q8SRZ6	Q8srz6 encephalito
26	30	83.3	389	17	Q8TQ40	Q8tq40 methanosarc
27	30	83.3	406	2	Q93CM9	Q93cm9 helicobacte
28	30	83.3	437	5	Q965E7	Q965e7 paramecium
29	30	83.3	437	5	Q965E6	Q965e6 paramecium
30	30	83.3	502	5	Q8T813	Q8t813 dictyosteli
31	30	83.3	586	3	Q9HG02	Q9hg02 emericella
32	30	83.3	950	12	Q66678	Q66678 equine herp
33	30	83.3	975	12	O39275	O39275 equine herp
34	30	83.3	980	12	Q66682	Q66682 equine herp
35	30	83.3	1315	3	Q9URX8	Q9urx8 schizosacch
36	29	80.6	208	2	Q51868	Q51868 pasteurella
37	29	80.6	233	2	Q51865	Q51865 pasteurella
38	29	80.6	263	2	Q9KIW5	Q9kiw5 thiobacillu
39	29	80.6	288	2	Q52345	Q52345 unclassifie
40	29	80.6	318	16	P95279	P95279 mycobacteri
41	29	80.6	321	5	Q9XWD4	Q9xwd4 caenorhabdi
42	29	80.6	415	2	O24782	O24782 pseudomonas
43	29	80.6	452	16	Q8UJH1	Q8ujh1 agrobacteri
44	29	80.6	478	2	Q54231	Q54231 streptococc
45	29	80.6	532	16	Q9K6D7	Q9k6d7 bacillus ha

ALIGNMENTS

RESULT 1
Q68981
ID Q68981 PRELIMINARY; PRT; 583 AA.
AC Q68981;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSV1 (Mutant strain tsB5), glycoprotein B (GB) gene (Fragment).
OS human herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10298;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84301857; PubMed=6089415;
RA Bzik D.J., Fox B.A., DeLuca N.A., Person S.;
RT "Nucleotide sequence of a region of the herpes simplex virus type 1 gB
RT glycoprotein gene: Mutations affecting rate of virus entry and cell
RT fusion.";
RL Virology 137:185-190(1984).
DR EMBL; K02720; AAA45777.1; -.
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B; 1.
DR ProDom; PD000693; Glycoprot_B; 1.
FT NON_TER 1 1
SQ SEQUENCE 583 AA; 64914 MW; 9F4997AFCC96E457 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 583;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSIEFARL 8
||||||||
Db 178 SSIEFARL 185

RESULT 2
Q9YW93
ID Q9YW93 PRELIMINARY; PRT; 887 AA.

```
AC   Q9YW93;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
GN   UL27.
OS   macropodid herpesvirus type 1 (parma wallaby herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   unclassified Herpesviridae.
OX   NCBI_TaxID=83441;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99171712; PubMed=10073704;
RA   Mahony T.J., Smith G.A., Thomson D.M.;
RT   "Macropodid herpesviruses 1 and 2 occupy unexpected molecular
RT   phylogenic positions within the Alphaherpesvirinae.";
RL   J. Gen. Virol. 80:433-436(1999).
DR   EMBL; AF061754; AAD11960.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   887 AA;  99739 MW;  0EE4FBB4E098F128 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 887;
  Best Local Similarity   100.0%;  Pred. No. 6.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     482 SSIEFARL 489


RESULT 3
Q65540
ID   Q65540      PRELIMINARY;      PRT;   893 AA.
AC   Q65540;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   GB glycoprotein.
GN   UL27.
OS   Baboon herpesvirus 2.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=36347;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X313;
RX   MEDLINE=95251550; PubMed=7733825;
RA   Eberle R., Black D.H., Lipper S., Hilliard J.K.;
RT   "Herpesvirus papio 2, an SA8-like alpha-herpesvirus of baboons.";
RL   Arch. Virol. 140:529-545(1995).
DR   EMBL; U14663; AAA85650.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   893 AA;  98520 MW;  06D9E973EE1712A1 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 893;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     487 SSIEFARL 494


RESULT 4
Q9YW92
ID   Q9YW92      PRELIMINARY;      PRT;   894 AA.
AC   Q9YW92;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
GN   UL27.
OS   macropodid herpesvirus type 2 (dorcopsis wallaby herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   unclassified Herpesviridae.
OX   NCBI_TaxID=83440;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99171712; PubMed=10073704;
RA   Mahony T.J., Smith G.A., Thomson D.M.;
RT   "Macropodid herpesviruses 1 and 2 occupy unexpected molecular
RT   phylogenic positions within the Alphaherpesvirinae.";
RL   J. Gen. Virol. 80:433-436(1999).
DR   EMBL; AF061755; AAD11961.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   894 AA;  100415 MW;  806C1DD45B72AECB CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 894;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SSIEFARL 8
           ||||||||
Db     485 SSIEFARL 492


RESULT 5
Q69464
ID   Q69464      PRELIMINARY;      PRT;   901 AA.
AC   Q69464;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B precursor.
GN   UL27.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CAM4B;
RA   Terhune S.S., Spear P.G.;
RT   "Variability of herpes simplex virus glycoproteins in clinical
RT   specimens.";
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 80-901 FROM N.A.
RC   STRAIN=CAM4B;
RX   MEDLINE=87112925; PubMed=3027364;
RA   Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RA   Burke R.L.;
RT   "Structure and expression of the herpes simplex virus type 2
RT   glycoprotein gB gene.";
RL   J. Virol. 61:326-335(1987).
DR   EMBL; U12172; AAB60545.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   VARIANT     46     46       Q -> R (IN REF. 2).
FT   VARIANT     57     57       R -> K (IN REF. 2).
FT   VARIANT    327    327       Q -> R (IN REF. 2).
SQ   SEQUENCE   901 AA;  99919 MW;  C1AE8BB4E000885E CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 SSIEFARL 8
          ||||||||
Db    493 SSIEFARL 500


RESULT 6
Q69465
ID   Q69465      PRELIMINARY;      PRT;   901 AA.
AC   Q69465;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B precursor.
GN   UL27.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MMA;
RA   Terhune S.S., Spear P.G.;
RT   "Variability of herpes simplex virus glycoproteins in clinical
RT   specimens.";
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 80-901 FROM N.A.
RC   STRAIN=MMA;
RX   MEDLINE=87112925; PubMed=3027364;
RA   Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RA   Burke R.L.;
RT   "Structure and expression of the herpes simplex virus type 2
RT   glycoprotein gB gene.";
RL   J. Virol. 61:326-335(1987).
DR   EMBL; U12174; AAB60547.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   VARIANT     57     57       R -> K (IN REF. 2).
FT   VARIANT     68     68       E -> K (IN REF. 2).
FT   VARIANT    327    327       Q -> R (IN REF. 2).
FT   VARIANT    393    393       E -> Q (IN REF. 2).
SQ   SEQUENCE   901 AA;  99949 MW;  6786F8DC747A4DF8 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 901;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    493 SSIEFARL 500


RESULT 7
Q69095
ID   Q69095      PRELIMINARY;      PRT;   902 AA.
AC   Q69095;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Glycoprotein B precursor.
GN   GB2.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=333;
RX   MEDLINE=88079667; PubMed=2446730;
RA   Zwaagstra J.C., Leung W.C.;
RT   "The nucleotide sequence of herpes simplex virus type 2 (333)
RT   glycoprotein gB2 and analysis of predicted antigenic sites.";
RL   Can. J. Microbiol. 33:879-887(1987).
DR   EMBL; M24771; AAA60540.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    902       GLYCOPROTEIN B.
SQ   SEQUENCE   902 AA;  100357 MW;  6232E753A63E6C73 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 902;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    494 SSIEFARL 501


RESULT 8
Q69076
ID   Q69076      PRELIMINARY;      PRT;   903 AA.
AC   Q69076;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
GN   GB-1.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10298;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88306231; PubMed=2841793;
RA   Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
RA   Buhk H.J.;
RT   "Conservation of a gene cluster including glycoprotein B in bovine
RT   herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL   Virology 165:388-405(1988).
DR   EMBL; M21629; AAA19496.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   903 AA;  100090 MW;  BA2D5611906307EE CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 903;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    498 SSIEFARL 505


RESULT 9
Q69526
ID   Q69526      PRELIMINARY;      PRT;   904 AA.
AC   Q69526;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Glycoprotein B.
GN   GB.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10298;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANG PATH;
```

```
RA   Holland T.C., Saharkhiz-Langroodi A.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U49121; AAA91805.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100291 MW;  D9C3C3B66FB1E033 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 10
Q9QLM8
ID   Q9QLM8      PRELIMINARY;      PRT;   904 AA.
AC   Q9QLM8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
GN   UL27.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10298;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HSZP;
RX   MEDLINE=20227371; PubMed=10766304;
RA   Kosovsky J., Vojvodova A., Oravcova I., Kudelova M., Matis J.,
RA   Rajcani J.;
RT   "Herpes simplex virus 1 (HSV-1) strain HSZP glycoprotein B gene :
RT   comparison of mutations among strains differing in virulence.";
RL   Virus Genes 20:27-33(2000).
DR   EMBL; AF097023; AAF04615.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100145 MW;  DF2C4D503E594D08 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 11
Q9IWU4
ID   Q9IWU4      PRELIMINARY;      PRT;   904 AA.
AC   Q9IWU4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Glycoprotein B.
GN   UL27.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10298;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=+GC;
RA   Ling J.-Y., Chen T.-M., Stroop W.G.;
RT   "A viral genetic element involved in a rabbit model of herpes simplex
RT   virus-induced epileptiform seizures.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF259899; AAF70301.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100185 MW;  9757F59FC8FEAD40 CRC64;

  Query Match             100.0%;  Score.36;  DB 12;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 12
Q9DXE4
ID   Q9DXE4      PRELIMINARY;      PRT;   904 AA.
AC   Q9DXE4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
OS   Herpes simplex virus (type 1 / strain KOS).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10306;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOS;
RX   MEDLINE=21066717; PubMed=11145912;
RA   Pertel P.E., Fridberg A., Parish M.L., Spear P.G.;
RT   "Cell fusion induced by herpes simplex virus glycoproteins gB, gD, and
RT   gH-gL requires a gD receptor but not necessarily heparan sulfate.";
RL   Virology 279:313-324(2001).
DR   EMBL; AF311740; AAG34116.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100294 MW;  21FD32F5E5F2AD48 CRC64;

  Query Match             100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    499 SSIEFARL 506


RESULT 13
O37453
ID   O37453      PRELIMINARY;      PRT;   904 AA.
AC   O37453;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Glycoprotein B.
GN   UL27.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BBKC;
RX   MEDLINE=98314525; PubMed=9652417;
RA   Terhune S.S., Coleman K.T., Sekulovich R., Burke R.L., Spear P.G.;
RT   "Limited variability of glycoprotein gene sequences and neutralizing
RT   targets in herpes simplex virus type 2 isolates and stability on
RT   passage in cell culture.";
RL   J. Infect. Dis. 178:8-15(1998).
```

```
DR   EMBL; AF021340; AAB72100.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100215 MW;  2170D008154B131F CRC64;

  Query Match           100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity  100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    496 SSIEFARL 503


RESULT 14
Q91C63
ID   Q91C63      PRELIMINARY;      PRT;   904 AA.
AC   Q91C63;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Glycoprotein B2.
GN   GB2.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G;
RA   Lee H.H.;
RT   "Sequencing and high level expression of the HSV-2 strain G
RT   glycoprotein B gene using a baculvirus HcNPV.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295528; AAK97852.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
SQ   SEQUENCE   904 AA;  100153 MW;  210DAAA5DD2475C6 CRC64;

  Query Match           100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity  100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    496 SSIEFARL 503


RESULT 15
Q89920
ID   Q89920      PRELIMINARY;      PRT;   904 AA.
AC   Q89920;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Glycoprotein B precursor.
GN   UL27.
OS   Herpes simplex virus (type 2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10310;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WTW1A;
RA   Terhune S.S., Spear P.G.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 80-904 FROM N.A.
RC   STRAIN=WTW1A;
RX   MEDLINE=87112925; PubMed=3027364;
RA   Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RA   Burke R.L.;
RT   "Structure and expression of the herpes simplex virus type 2
RT   glycoprotein gB gene.";
RL   J. Virol. 61:326-335(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WTW1A;
RA   Spear P.G.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U12175; AAB60548.1; -.
DR   EMBL; U12173; AAB60546.1; -.
DR   InterPro; IPR000234; Glycoprot_B.
DR   Pfam; PF00606; Glycoprotein_B; 1.
DR   ProDom; PD000693; Glycoprot_B; 1.
KW   Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   VARIANT     396    396       E -> Q (IN REF. 2).
SQ   SEQUENCE   904 AA;  100187 MW;  A8B36F74FBE72139 CRC64;

  Query Match           100.0%;  Score 36;  DB 12;  Length 904;
  Best Local Similarity  100.0%;  Pred. No. 6.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SSIEFARL 8
          ||||||||
Db    496 SSIEFARL 503


Search completed: May 22, 2003, 12:09:37
Job time : 28 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:16 ; Search time 31 Seconds
(without alignments)
34.387 Million cell updates/sec

Title: US-09-719-494-12
Perfect score: 44
Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	8	21	AAY67357	Melanoma glycoprot
2	37	84.1	1172	23	AAU74786	Mouse thrombospond
3	37	84.1	1172	23	ABB72334	Rat protein isolat
4	36	81.8	1199	19	AAW47206	Bos taurus tubulin
5	36	81.8	1371	22	ABB52745	Escherichia coli p
6	35	79.5	513	22	ABG07347	Novel human diagno
7	35	79.5	571	21	AAB43586	Human cancer assoc
8	35	79.5	571	22	AAG75130	Human colon cancer
9	35	79.5	1172	21	AAB19677	Human thrombospond
10	35	79.5	1172	21	AAB00043	Human thrombospond
11	35	79.5	1172	23	AAU76902	Human Thrombospond
12	35	79.5	1172	23	AAU74788	Human thrombospond
13	33	75.0	91	21	AAY70091	Porcine endogenous
14	33	75.0	705	22	ABG29447	Novel human diagno
15	33	75.0	1766	22	ABB58631	Drosophila melanog
16	32	72.7	8	21	AAY67360	Melanoma glycoprot
17	32	72.7	9	21	AAB03467	Human tyrosine-rel
18	32	72.7	9	22	AAU26665	Human Leukocyte An
19	32	72.7	9	22	AAU26998	Human Leukocyte An
20	32	72.7	10	22	AAU26707	Human Leukocyte An
21	32	72.7	10	22	AAU27040	Human Leukocyte An
22	32	72.7	30	23	AAU84873	Human Trp-1 segmen
23	32	72.7	30	23	AAU84938	Human Trp2 segment
24	32	72.7	79	22	AAO06853	Human polypeptide
25	32	72.7	134	23	ABB10041	Tyrosinase 15. Un
26	32	72.7	203	22	AAU23172	Novel human enzyme
27	32	72.7	248	23	AAU84805	Human TRP-1 consen
28	32	72.7	249	12	AAR14258	gp75 peptide and f
29	32	72.7	253	22	AAE10135	Streptomyces nours
30	32	72.7	306	17	AAW05511	HCMV Toledo strain
31	32	72.7	390	22	AAM78765	Human protein SEQ
32	32	72.7	474	22	AAM79749	Human protein SEQ
33	32	72.7	486	22	AAB93541	Human protein sequ
34	32	72.7	491	22	ABG07344	Novel human diagno
35	32	72.7	498	22	ABG09680	Novel human diagno
36	32	72.7	517	20	AAY42635	Murine tyrosinase-
37	32	72.7	517	20	AAY31981	Mouse tyrosinase-r
38	32	72.7	517	22	AAB74679	Human protease and
39	32	72.7	517	23	AAU76663	Mouse tyrosine-rel
40	32	72.7	517	23	AAE14345	Human protease PRT
41	32	72.7	517	23	AAU82718	Amino acid sequenc
42	32	72.7	519	18	AAW30826	The novel tyrosina
43	32	72.7	519	20	AAY42636	Human tyrosinase-r
44	32	72.7	519	20	AAY31982	Human tyrosinase-r
45	32	72.7	519	22	AAB86041	Human Trp2 protein

ALIGNMENTS

RESULT 1
AAY67357
ID AAY67357 standard; peptide; 8 AA.
XX
AC AAY67357;
XX
DT 25-APR-2000 (first entry)
XX
DE Melanoma glycoprotein 75 peptide TAY, used as a therapeutic antigen.
XX
KW Therapeutic antigen; cytotoxic T lymphocyte; CTL; CTL immune response;
KW cellular immune response induction method; vaccine; human; tumour;
KW melanoma glycoprotein 75.
XX
OS Homo sapiens.
XX
PN WO9963945-A2.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13146.
XX
PR 12-JUN-1998; 98US-0089055.
PR 30-OCT-1998; 98US-0106339.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PI Nikolic-Zugic J, Dyall R, Houghton AN;
XX
DR WPI; 2000-126432/11.
XX
PT Induction of a cellular immune response to a weakly immunogenic
PT protein, used to target and kill tumour cells -

XX
PS Claim 14; Page 24; 44pp; English.
XX
CC This sequence represents a melanoma glycoprotein 75 peptide used as a
CC therapeutic antigen in the method of the invention. The invention relates
CC to a method for inducing a cytotoxic T lymphocyte (CTL) immune response
CC to non/weakly-immunogenic proteins which are expressed on tumour cells.
CC The method for inducing a cellular immune response to a non-immunogenic
CC or weakly immunogenic target peptide expressed on tumour cells of a
CC mammalian subject comprises administering antigen to induce a cellular
CC immune response to the target peptide. The antigen comprises an
CC immunogenic portion having a major histocompatibility complex (MHC)
CC binding domain which binds to the MHC and an immune recognition domain
CC which is recognized by T-cells. The antigen is derived from the target
CC peptide such that the MHC-binding portion binds to MHC with a greater
CC affinity than the target peptide without material alteration of the
CC immune recognition portion. The methods are used for inducing a cellular
CC immune response to a non-immunogenic or weakly immunogenic target peptide
CC expressed on tumour cells of a mammalian subject. The antigens and
CC immunogens of the invention, as well as polynucleotides encoding them,
CC are used in vaccine compositions against tumour cells.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 44; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAYRYHLL 8
||||||||
Db 1 TAYRYHLL 8

RESULT 2
AAU74786
ID AAU74786 standard; Protein; 1172 AA.
XX
AC AAU74786;
XX
DT 09-APR-2002 (first entry)
XX
DE Mouse thrombospondin-2 (TSP-2).
XX
KW Thrombospondin-2; TSP-2; cytostatic; angiogenesis; vasotropic;
KW vulnerary; neovascularisation; cell proliferation inhibitor; cancer;
KW solid tumour; haemangioma; acoustic neuromas; neurofibroma; trachoma;
KW pyogenic granulomas; rheumatoid arthritis; ocular angiogenic disease;
KW retinopathy; psoriasis; macular degeneration; corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; rebeosis; angiofibroma;
KW Osler-Webber syndrome; myocardial angiogenesis; haemophiliac joints;
KW plaque neovascularisation; telangiectasia; wound granulation;
KW mouse; apoptosis.
XX
OS Mus sp.
XX
PN WO200191781-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US17250.
XX
PR 26-MAY-2000; 2000US-207994P.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lawler JW;
XX
DR WPI; 2002-106273/14.
XX
PT Composition useful for treatment of cancer comprises cDNA encoding
PT amino acids of human thrombospondin-1 or its conservative variant and a
PT carrier -
XX
PS Example 2; Fig 10; 54pp; English.
XX
CC The invention describes a composition comprising cDNA encoding fragments
CC of human thrombospondin-1 (TSP -1), a type 1 repeat polypeptide and
CC potent inhibitor of tumour growth and angiogenesis. The composition is
CC useful for killing cancerous cells (preferably tumour); for reducing
CC volume or inhibiting growth of a tumour (inhibiting neovascularisation in
CC the tumour); for decreasing proliferation of tumour cells; in the
CC treatment of diseases and conditions associated with angiogenic activity
CC or misregulated growth and angiogenesis-mediated diseases such as cancer,
CC solid tumour, tumour metastasis, benign tumour, (e.g. haemangioma,
CC acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas),
CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g.
CC diabetic retinopathy, retinopathy of prematurity, macular degeneration,
CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasias,
CC rebeosis), Osler-Webber syndrome, myocardial angiogenesis,
CC telangiectasia, plaque neovascularisation, haemophiliac joints,
CC angiofibroma or wound granulation. The composition induces apoptosis and
CC inhibits neovascularisation in the tumour cells. This amino acid sequence
CC represents mouse thrombospondin-2 (TSP-2), on which the recombinant
CC protein (see AAU74796) of the invention is based.
XX
SQ Sequence 1172 AA;

Query Match 84.1%; Score 37; DB 23; Length 1172;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAYRYHLL 8
||||:||:
Db 1108 TAYRWHLI 1115

RESULT 3
ABB72334
ID ABB72334 standard; Protein; 1172 AA.
XX
AC ABB72334;
XX
DT 04-APR-2002 (first entry)
XX
DE Rat protein isolated from skin cells SEQ ID NO: 658.
XX
KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW developmental defect; inflammatory disease; dermatological; vulnerary;
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX
OS Rattus sp.
XX
PN WO200190357-A1.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-NZ00099.
XX
PR 24-MAY-2000; 2000US-206650P.
PR 25-JUL-2000; 2000US-221232P.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;
XX
DR WPI; 2002-122020/16.
DR N-PSDB; ABL35019.
XX
PT New polynucleotides and polypeptides encoded by the polynucleotides
PT isolated from skin cells, useful for treating skin wounds, cancers,
PT growth and developmental defects, inflammatory diseases, or for
PT modulating immune responses -
XX
PS Claim 4; Page 413-416; 466pp; English.
XX
CC The present invention provides the protein and coding sequences of cDNAs

```
CC   isolated from human, murine and rat skin cell libraries. The sequences
CC   can be used in the development of therapeutic agents useful in the
CC   treatment of skin diseases, including skin wounds, cancer, growth
CC   defects, developmental defects and inflammatory diseases. The proteins
CC   have important roles in the induction of hair growth, cell proliferation
CC   and cell-cell interaction, in maintaining tissue integrity, in wound
CC   healing and in modulating immune responses. The present sequence is a
CC   polypeptide of the invention.
XX
SQ   Sequence   1172 AA;

  Query Match            84.1%;  Score 37;  DB 23;  Length 1172;
  Best Local Similarity  75.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYHLL 8
          ||||:||:
Db   1108 TAYRWHLI 1115


RESULT 4
AAW47206
ID   AAW47206 standard; Protein; 1199 AA.
XX
AC   AAW47206;
XX
DT   21-JUL-1998  (first entry)
XX
DE   Bos taurus tubulin-folding cofactor D.
XX
KW   Tubulin folding; cofactor; alpha-tubulin; beta-tubulin; unfolded;
KW   folded; treatment; hyper-proliferative diseases; cancer; gout.
XX
OS   Bos taurus.
XX
PN   WO9804587-A1.
XX
PD   05-FEB-1998.
XX
PF   25-JUL-1997;   97WO-US14076.
XX
PR   25-JUL-1996;   96US-0023089.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Cowan NJ;
XX
DR   WPI; 1998-130618/12.
DR   N-PSDB; AAV17086.
XX
PT   New isolated cofactor(s) for tubulin folding - are useful as targets
PT   for identifying agents which interfere with folding in the treatment
PT   of hyper-proliferative diseases such as cancer
XX
PS   Claim 3; Pages 48-52; 87pp; English.
XX
CC   The sequence is that of bovine tubulin-folding cofactor D.
CC   It may be useful as a target for interfering with the
CC   production of productively folded alpha- and beta-tubulins.
CC   Since tubulin function is essential for cell division and
CC   proliferation, agents which interfere with tubulin function
CC   can serve as useful antiproliferative compounds. Such interfering
CC   agents have potential utility as agents for the treatment of
CC   hyperproliferative diseases such as cancer and the treatment
CC   of gout.
XX
SQ   Sequence   1199 AA;

  Query Match            81.8%;  Score 36;  DB 19;  Length 1199;
  Best Local Similarity  85.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 AYRYHLL 8
            |||||:|
Db    985 AYRYHVL 991


RESULT 5
ABB52745
ID   ABB52745 standard; Protein; 1371 AA.
XX
AC   ABB52745;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 891.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1371 AA;

  Query Match            81.8%;  Score 36;  DB 22;  Length 1371;
  Best Local Similarity  85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYHL 7
            | |||||
Db   1226 TGYRYHL 1232


RESULT 6
ABG07347
ID   ABG07347 standard; Protein; 513 AA.
XX
AC   ABG07347;
XX
```

DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #7338.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS71534.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 37706; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 513 AA;

Query Match 79.5%; Score 35; DB 22; Length 513;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YRYHLL 8
||||||
Db 493 YRYHLL 498

RESULT 7
AAB43586
ID AAB43586 standard; Protein; 571 AA.
XX
AC AAB43586;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1031.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR N-PSDB; AAC77795.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 11; Page 1616-1619; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 571 AA;

Query Match 79.5%; Score 35; DB 21; Length 571;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAYRYHL 7
||||:||
Db 507 TAYRWHL 513

RESULT 8
AAG75130
ID AAG75130 standard; Protein; 571 AA.
XX
AC AAG75130;

```
XX
DT  03-SEP-2001  (first entry)
XX
DE  Human colon cancer antigen protein SEQ ID NO:5894.
XX
KW  Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW  colorectal carcinoma; chromosome 6.
XX
OS  Homo sapiens.
XX
PN  WO200122920-A2.
XX
PD  05-APR-2001.
XX
PF  28-SEP-2000; 2000WO-US26524.
XX
PR  29-SEP-1999;   99US-0157137.
PR  03-NOV-1999;   99US-0163280.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR  WPI; 2001-235357/24.
DR  N-PSDB; AAH34535.
XX
PT  Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT  useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS  Claim 11; Page 7409-7411; 9803pp; English.
XX
CC  AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC  cancer-associated nucleic acid molecules (N) and proteins (P), where
CC  the proteins are collectively known as colon cancer antigens. The colon
CC  cancer antigens have cytostatic activity and can be used in gene
CC  therapy and vaccine production. N and P may be used in the prevention,
CC  diagnosis and treatment of diseases associated with inappropriate P
CC  expression. For example, N and P may be used to treat disorders
CC  associated with decreased expression by rectifying mutations or deletions
CC  in a patient's genome that affect the activity of P by expressing
CC  inactive proteins or to supplement the patients own production of P.
CC  Additionally, N may be used to produce the colon cancer-associated Ps,
CC  by inserting the nucleic acids into a host cell and culturing the cell
CC  to express the proteins. N and P can be used in the prevention, diagnosis
CC  and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC  and AAB77789 represent sequences used in the exemplification of the
CC  present invention.
CC  N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC  missing at time of publication, meaning no sequences are present for
CC  SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ  Sequence   571 AA;

  Query Match             79.5%;  Score 35;  DB 22;  Length 571;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db     507 TAYRWHL 513


RESULT 9
AAB19677
ID  AAB19677 standard; Protein; 1172 AA.
XX
AC  AAB19677;
XX
DT  05-FEB-2001  (first entry)
XX
DE  Human thrombospondin-2.
XX
KW  Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
KW  melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;
KW  prostate cancer; psoriasis; rosecea dermatosis; antitumour;
KW  therapy.
XX
OS  Homo sapiens.
XX
FH  Key             Location/Qualifiers
FT  Region          382..429
FT                  /note= "type 1 repeat"
FT  Region          384..390
FT                  /note= "WSPWAEW sequence involved in antiangiogenic
FT                        activity"
FT  Region          438..490
FT                  /note= "type 1 repeat"
FT  Region          495..547
FT                  /note= "type 1 repeat"
XX
PN  WO200057899-A1.
XX
PD  05-OCT-2000.
XX
PF  24-MAR-2000; 2000WO-US07835.
XX
PR  31-MAR-1999;   99US-0127221.
XX
PA  (GEHO ) GEN HOSPITAL CORP.
XX
PI  Detmar M,  Streit M;
XX
DR  WPI; 2000-656131/63.
DR  N-PSDB; AAA88669.
XX
PT  Treating a disorder characterized by unwanted cell proliferation e.g.
PT  precancerous, cancerous or neoplastic cells or presence of tumor
PT  preferably of skin or prostate, comprises increasing thrombospondin-2
PT  activity  -
XX
PS  Disclosure; Fig 2; 73pp; English.
XX
CC  The present sequence is that of human thrombospondin-2 (TSP-2).
CC  The invention is based on the discovery that overexpression of
CC  TSP-2 decreases tumor size in vivo, and features methods for
CC  modulating unwanted angiogenesis and tumour growth.  Treatment of
CC  unwanted cell proliferation or angiogenesis involves increasing
CC  TSP-2 activity.  This is achieved by administering an agent which
CC  increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2
CC  derived polypeptide or retro-inverso peptide, a nucleic acid
CC  encoding TSP-2, an agonist of TSP-2, or an agent that increases
CC  TSP-2 gene expression.  The TSP-2 polypeptide may include at least
CC  one type 1 repeat such as the WSPWAEW peptide (see AAB19683),
CC  which is involved in the antiangiogenic activity of TSP-2.  The
CC  method is used to treat a disorder characterised by pre-cancerous,
CC  cancerous or neoplastic cells, or the presence of a tumour, or a
CC  disorder that affects epithelial tissues resulting in unwanted
CC  skin cell proliferation.  Such disorders include malignant
CC  melanoma, prostate cancer, squamous cell carcinoma, aged skin,
CC  rosecea dermatosis, psoriasis, and skin damage caused by
CC  photoradiation (all claimed).  Evaluating the presence of TSP-2
CC  nucleic acid or protein is useful for diagnosing a subject at risk
CC  of unwanted cell proliferation or angiogenesis.  Methods are also
CC  provided of identifying compounds that modulate TSP-2 activity.
XX
SQ  Sequence   1172 AA;

  Query Match             79.5%;  Score 35;  DB 21;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1108 TAYRWHL 1114
```

```
RESULT 10
AAB00043
ID   AAB00043 standard; Protein; 1172 AA.
XX
AC   AAB00043;
XX
DT   08-NOV-2000  (first entry)
XX
DE   Human thrombospondon-2 (TSP-2).
XX
KW   TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW   thrombospondin; angiogenesis; tumour; treatment; cancer;
KW   arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KW   glaucoma.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          381..436
FT                   /label= Type 1 repeat region
FT   Region          437..493
FT                   /label= Type 1 repeat region
FT   Region          494..550
FT                   /label= Type 1 repeat region
XX
PN   WO200044908-A2.
XX
PD   03-AUG-2000.
XX
PF   01-FEB-2000; 2000WO-US02482.
XX
PR   01-FEB-1999;   99US-0118053.
XX
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Lawler JW;
XX
DR   WPI; 2000-514823/46.
XX
PT   Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT   matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
PT   inhibiting angiogenesis and treating diseases such as cancer
XX
PS   Disclosure; Fig 2; 40pp; English.
XX
CC   New nucleic acids are described which encode a protein comprising
CC   the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC   but not the TGF (transforming growth factor)-beta activation region
CC   of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC   the second and third type-1 repeats and the COMP (cartilage
CC   oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC   produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC   caused inhibition of the growth of tumours in mice models.
CC   Thus the nucleic acids and proteins may be useful for treating
CC   angiogenesis related diseases such as cancer (by reducing the rate of
CC   growth and size of tumours), arthritis, psoriasis, diabetic
CC   retinopathy, corneal graft rejection, and glaucoma. They may also be
CC   used for treating human immunodeficiency virus (HIV) infection.
CC   Anti-angiogenic therapy has little toxicity, does not require the
CC   therapeutic agent to enter tumour cells or cross the blood-brain
CC   barrier, controls tumour growth independently of growth of
CC   tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ   Sequence   1172 AA;

  Query Match             79.5%;  Score 35;  DB 21;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1108 TAYRWHL 1114
```

```
RESULT 11
AAU76902
ID   AAU76902 standard; Protein; 1172 AA.
XX
AC   AAU76902;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human Thrombospondin 2 protein.
XX
KW   Human; Osteopontin 1; OPN1; thrombospondin 2;
KW   wound response; foreign body response; cut; abrasion; burn;
KW   vulnerary.
XX
OS   Homo sapiens.
XX
PN   WO200209735-A2.
XX
PD   07-FEB-2002.
XX
PF   31-JUL-2001; 2001WO-US24147.
XX
PR   01-AUG-2000; 2000US-222071P.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Bornstein P,  Kyriakides T,  Ratner B,  Giachelli C,  Martinson L;
PI   Scatena M;
XX
DR   WPI; 2002-217098/27.
DR   P-PSDB; ABK10277.
XX
PT   Modulating the amount or biological activity of thrombospondin 2 or
PT   osteopontin in an animal for modulating a wound response, comprises
PT   introducing osteopontin or thrombospondin 2 antagonist into an animal -
XX
PS   Disclosure; Page 50-54; 54pp; English.
XX
CC   This invention relates to a method for modulating the amount or
CC   biological activity of thrombospondin 2 or osteopontin in an animal. The
CC   method involves introducing into the animal an osteopontin or
CC   thrombospondin 2 antagonist. Using the methods of the invention the
CC   amount or biological activity of thrombospondin 2 or osteopontin
CC   protein may be modulated. The method of the invention is useful for
CC   modulating the amount or biological activity of thrombospondin 2 or
CC   osteopontin in an animal which exhibits a wound response or a foreign
CC   body response, where the method can be used to improves the wound
CC   response or reduces the foreign body response. The method is
CC   useful for modulating a wound response or for reducing a foreign body
CC   response in an animal and is also useful for improving the wound
CC   response, such as at the site of a cut, abrasion or burn. The
CC   present sequence represents the human thrombospondin 2 protein used
CC   in the method of the invention to modulate wound healing.
XX
SQ   Sequence   1172 AA;

  Query Match             79.5%;  Score 35;  DB 23;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1108 TAYRWHL 1114


RESULT 12
AAU74788
ID   AAU74788 standard; Protein; 1172 AA.
XX
AC   AAU74788;
XX
DT   09-APR-2002  (first entry)
```

```
XX
DE   Human thrombospondin-2 (TSP-2).
XX
KW   Thrombospondin-2; TSP-2; cytostatic; angiogenesis; vasotropic;
KW   vulnerary; neovascularisation; cell proliferation inhibitor; cancer;
KW   solid tumour; haemangioma; acoustic neuromas; neurofibroma; trachoma;
KW   pyogenic granulomas; rheumatoid arthritis; ocular angiogenic disease;
KW   retinopathy; psoriasis; macular degeneration; corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; rebeosis; angiofibroma;
KW   Osler-Webber syndrome; myocardial angiogenesis; haemophiliac joints;
KW   plaque neovascularisation; telangiectasia; wound granulation; apoptosis.
XX
OS   Homo sapiens.
XX
PN   WO200191781-A2.
XX
PD   06-DEC-2001.
XX
PF   25-MAY-2001; 2001WO-US17250.
XX
PR   26-MAY-2000; 2000US-207994P.
XX
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Lawler JW;
XX
DR   WPI; 2002-106273/14.
XX
PT   Composition useful for treatment of cancer comprises cDNA encoding
PT   amino acids of human thrombospondin-1 or its conservative variant and a
PT   carrier -
XX
PS   Disclosure; Fig 11; 54pp; English.
XX
CC   The invention describes a composition comprising cDNA encoding fragments
CC   of human thrombospondin-1 (TSP -1), a type 1 repeat polypeptide and
CC   potent inhibitor of tumour growth and angiogenesis. The composition is
CC   useful for killing cancerous cells (preferably tumour); for reducing
CC   volume or inhibiting growth of a tumour (inhibiting neovascularisation in
CC   the tumour); for decreasing proliferation of tumour cells; in the
CC   treatment of diseases and conditions associated with angiogenic activity
CC   or misregulated growth and angiogenesis-mediated diseases such as cancer,
CC   solid tumour, tumour metastasis, benign tumour, (e.g. haemangioma,
CC   acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas),
CC   rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g.
CC   diabetic retinopathy, retinopathy of prematurity, macular degeneration,
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasias,
CC   rebeosis), Osler-Webber syndrome, myocardial angiogenesis,
CC   telangiectasia, plaque neovascularisation, haemophiliac joints,
CC   angiofibroma or wound granulation. The composition induces apoptosis and
CC   inhibits neovascularisation in the tumour cells. This amino acid sequence
CC   represents human thrombospondin-2 (TSP-2), on which the recombinant
CC   protein (shown in AAU74789) of the invention is based.
XX
SQ   Sequence   1172 AA;

  Query Match             79.5%;  Score 35;  DB 23;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1108 TAYRWHL 1114


RESULT 13
AAY70091
ID   AAY70091 standard; Protein; 91 AA.
XX
AC   AAY70091;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Porcine endogenous retrovirus-D env polypeptide.
XX
KW   Porcine endogenous retrovirus; PERV-D; virucide; prevention; vaccine;
KW   pig; diagnosis; infection; xenotransplantation; antibody; env region.
XX
OS   Sus scrofa.
XX
PN   WO200011187-A1.
XX
PD   02-MAR-2000.
XX
PF   18-AUG-1999;   99WO-US19053.
XX
PR   18-AUG-1998;   98US-0097015.
XX
PA   (BIOT-) BIO TRANSPLANT INC.
XX
PI   Banerjee PT,  Patience C,  Andersson GK;
XX
DR   WPI; 2000-224704/19.
DR   N-PSDB; AAZ51065.
XX
PT   Porcine retroviral PERV-D polypeptides for diagnosing porcine
PT   retroviral infections in humans after xenotransplantation -
XX
PS   Claim 9; Fig 2; 119pp; English.
XX
CC   The present amino acid sequence is the porcine endogenous retrovirus
CC   env polypeptide-D (PERV-D). It is isolated from the porcine kidney, PK15
CC   cell line (ATCC No.CCL-33). PERV-D polynucleotide has 79% homology to a
CC   portion of PERV-C and has virucidal activity. PERV-D sequence is useful
CC   for prevention or diagnosis of infection of human tissues by porcine
CC   retroviruses after xenotransplantation procedures. PERV-D polypeptide
CC   may be used to produce specific antibodies, that can be administered as
CC   vaccines to create passive immunity. The DNA can be used as a
CC   hybridisation probe or primer for isolation purposes.
XX
SQ   Sequence   91 AA;

  Query Match             75.0%;  Score 33;  DB 21;  Length 91;
  Best Local Similarity   83.3%;  Pred. No. 40;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYH 6
             |:||||
Db      23 TSYRYH 28


RESULT 14
ABG29447
ID   ABG29447 standard; Protein; 705 AA.
XX
AC   ABG29447;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #29438.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
```

XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS93634.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 59806; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 705 AA;

Query Match 75.0%; Score 33; DB 22; Length 705;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAYRYHL 7
|::||||
Db 542 TSFRYHL 548

RESULT 15
ABB58631
ID ABB58631 standard; Protein; 1766 AA.
XX
AC ABB58631;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2685.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02734.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 2685; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1766 AA;

Query Match 75.0%; Score 33; DB 22; Length 1766;
Best Local Similarity 85.7%; Pred. No. 9.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AYRYHLL 8
|| ||||
Db 1399 AYTYHLL 1405

Search completed: May 22, 2003, 12:08:19
Job time : 32 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:06:12 ; Search time 14 Seconds
(without alignments)
54.934 Million cell updates/sec

Title: US-09-719-494-12
Perfect score: 44
Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	84.1	1172	2	A42587	thrombospondin 2 p
2	37	84.1	1178	1	A39804	thrombospondin pre
3	36	81.8	145	2	F75337	transposase - Dein
4	36	81.8	1199	2	T18522	tubulin-folding co
5	35	79.5	323	2	F98303	probable transcrip
6	35	79.5	323	2	AG2979	transcription regu
7	35	79.5	944	2	T47246	chitin synthase (E
8	35	79.5	1172	1	TSHUP2	thrombospondin 2 p
9	34	77.3	162	2	A70489	conserved hypothet
10	33	75.0	72	2	T17937	hypothetical prote
11	33	75.0	173	2	D72586	hypothetical prote
12	33	75.0	321	2	G70415	nucleotide sugar e
13	33	75.0	374	2	H81399	2-oxoglutarate-fer
14	32	72.7	150	2	S27613	hypothetical prote
15	32	72.7	172	2	AC2548	hypothetical prote
16	32	72.7	214	2	S44706	opacity protein op
17	32	72.7	234	2	S36329	opacity protein op
18	32	72.7	234	2	S36342	opacity protein op
19	32	72.7	234	2	S36341	opacity protein op
20	32	72.7	235	2	S44707	opacity protein op
21	32	72.7	248	2	PL0038	opacity protein D
22	32	72.7	254	2	S20043	opacity protein B
23	32	72.7	257	2	S16614	opacity protein op
24	32	72.7	258	2	S08514	opacity protein-re
25	32	72.7	268	1	KONH2C	opacity protein P.
26	32	72.7	283	2	S72343	opacity protein op
27	32	72.7	294	2	H83490	hypothetical prote
28	32	72.7	338	2	S16613	opacity protein op
29	32	72.7	352	2	I51282	TRP-1 - axolotl (f
30	32	72.7	366	2	T12548	hypothetical prote
31	32	72.7	414	2	S18962	FBF15 protein - St
32	32	72.7	517	1	S19243	tyrosinase-related
33	32	72.7	519	1	YRHUR2	dopachrome isomera
34	32	72.7	522	2	I51245	tyrosinase related
35	32	72.7	537	1	YRHUB6	tyrosinase-related
36	32	72.7	537	1	YRMSB6	tyrosinase-related
37	32	72.7	550	2	AI0807	probable decarboxy
38	32	72.7	679	2	T00636	hypothetical prote
39	32	72.7	1050	2	T43482	hypothetical prote
40	32	72.7	3951	1	VFIHB1	F1 protein - avian
41	31	70.5	72	2	S31028	gene 83 protein -
42	31	70.5	121	1	PC4024	phospholipase A2 h
43	31	70.5	129	2	C87219	hypothetical prote
44	31	70.5	141	2	A70556	probable mutator M
45	31	70.5	190	2	S18680	DNA-invertase - Es

ALIGNMENTS

RESULT 1
A42587
thrombospondin 2 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C;Accession: A42587; A39851
R;Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A;Title: Characterization of mouse thrombospondin 2 sequence and expression during
A;Reference number: A42587; MUID:92147683; PMID:1371115
A;Accession: A42587
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1172 <LAH>
A;Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A;Note: sequence extracted from NCBI backbone (NCBIP:81502)
R;Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V
J. Biol. Chem. 266, 12821-12824, 1991
A;Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome
A;Reference number: A39851; MUID:91302287; PMID:1712771
A;Accession: A39851
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-873 <BOR>
A;Cross-references: GB:M64866; NID:g201994; PIDN:AAA40432.1; PID:g201995
C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homolog
C;Keywords: calcium binding; glycoprotein
F;319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F;380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F;436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F;493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F;553-588/Domain: EGF homology <EGF1>
F;652-691/Domain: EGF homology <EGF>

Query Match 84.1%; Score 37; DB 2; Length 1172;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 TAYRYHLL 8
             ||||:||:
Db    1108 TAYRWHLI 1115
```

RESULT 2
A39804
thrombospondin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A39804
R;Lawler, J.; Duquette, M.; Ferro, P.
J. Biol. Chem. 266, 8039-8043, 1991
A;Title: Cloning and sequencing of chicken thrombospondin.

```
A;Reference number: A39804; MUID:91217026; PMID:2022631
A;Accession: A39804
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1178 <LAW>
A;Cross-references: GB:M60853; NID:g212763; PIDN:AAA51437.1; PID:g212764
C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vo
F;325-383/Domain: von Willebrand factor type C repeat homology <VWC>
F;386-437/Domain: thrombospondin type 1 repeat homology <THR1>
F;442-498/Domain: thrombospondin type 1 repeat homology <THR2>
F;499-555/Domain: thrombospondin type 1 repeat homology <THR3>
F;658-697/Domain: EGF homology <EGF>

  Query Match             84.1%;  Score 37;  DB 1;  Length 1178;
  Best Local Similarity   75.0%;  Pred. No. 28;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
            ||||:||:
Db    1114 TAYRWHLI 1121


RESULT 3
F75337
transposase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75337
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75337
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-145 <WHI>
A;Cross-references: GB:AE002032; GB:AE000513; NID:g6459715; PIDN:AAF11477.1; PID:g645970
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1927
A;Map position: 1
C;Superfamily: conserved hypothetical protein 4 (insertion sequence ISH1.8)

  Query Match             81.8%;  Score 36;  DB 2;  Length 145;
  Best Local Similarity   75.0%;  Pred. No. 5;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
            | ||||:|
Db      22 TKYRYHVL 29


RESULT 4
T18522
tubulin-folding cofactor D - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
C;Accession: T18522
R;Tian, G.; Huang, Y.; Rommelaere, H.; Vandekerckhove, J.; Ampe, C.; Cowan, N.J.
Cell 86, 287-296, 1996
A;Title: Pathway leading to correctly folded beta-tubulin.
A;Reference number: Z18945; MUID:96319731; PMID:8706133
A;Accession: T18522
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1199 <TIA>
A;Cross-references: EMBL:U61233; NID:g1465769; PID:g1465770; PIDN:AAB17537.1
C;Function:
A;Description: cofactors A and D are involved in capturing and stabilizing beta-tubulin
C;Superfamily: Arabidopsis thaliana beta-tubulin cofactor-like protein

  Query Match             81.8%;  Score 36;  DB 2;  Length 1199;
  Best Local Similarity   85.7%;  Pred. No. 46;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 AYRYHLL 8
            |||||:|
Db     985 AYRYHVL 991


RESULT 5
F98303
probable transcription regulator PA1184 [imported] - Agrobacterium tumefaciens (stra
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: F98303
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldb
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: F98303
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89952.1; PID:g15159911; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_2774
A;Map position: linear chromosome
C;Superfamily: regulatory protein ampR

  Query Match             79.5%;  Score 35;  DB 2;  Length 323;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
            || |||||
Db     287 TASRYHLL 294


RESULT 6
AG2979
transcription regulator, LysR family Atu3440 [imported] - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AG2979
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McC
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG2979
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44253.1; PID:g17741837; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3440
A;Map position: linear chromosome
C;Superfamily: regulatory protein ampR

  Query Match             79.5%;  Score 35;  DB 2;  Length 323;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
            || |||||
Db     287 TASRYHLL 294
```

```
RESULT 7
T47246
chitin synthase (EC 2.4.1.16) 2 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C;Accession: T47246
R;Beth Din, A.; Yarden, O.
Microbiology 140, 2189-2197, 1994
A;Title: The Neurospora crassa chs-2 gene encodes a non-essential chitin synthase.
A;Reference number: Z24422; MUID:95039879; PMID:7952169
A;Accession: T47246
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-944 <BET>
A;Cross-references: EMBL:X77782; NID:g558230; PIDN:CAA54816.1; PID:g558604
C;Genetics:
A;Gene: chs-2
A;Map position: IVR
A;Introns: 542/1
C;Superfamily: chitin synthase chsA
C;Keywords: glycosyltransferase; hexosyltransferase

  Query Match             79.5%;  Score 35;  DB 2;  Length 944;
  Best Local Similarity   75.0%;  Pred. No. 56;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYHLL 8
          :|||||.|
Db    468 SAYRYHAL 475


RESULT 8
TSHUP2
thrombospondin 2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C;Accession: A47379; A42173
R;LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993
A;Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: pote
A;Reference number: A47379; MUID:94010892; PMID:8406456
A;Accession: A47379
A;Molecule type: mRNA
A;Residues: 1-1172 <LAB>
A;Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506
R;LaBell, T.L.; Milewicz, D.J.; Disteche, C.M.; Byers, P.H.
Genomics 12, 421-429, 1992
A;Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expression c
A;Reference number: A42173; MUID:92217961; PMID:1559694
A;Accession: A42173
A;Molecule type: mRNA
A;Residues: 560-1172 <LA2>
A;Cross-references: GB:M81339
A;Experimental source: fibroblast
A;Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBIP:95096)
C;Genetics:
A;Gene: GDB:THBS2; TSP2
A;Cross-references: GDB:128789; OMIM:188061
A;Map position: 6q27-6q27
C;Complex: homotrimer, disulfide linked
C;Function:
A;Description: participates in cell migration and adhesion, and in platelet aggregation
C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vo
C;Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; trimer
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1172/Product: thrombospondin 2 #status predicted <MAT>
F;319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F;380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F;436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F;493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F;553-588/Domain: EGF homology <EGF1>
F;652-691/Domain: EGF homology <EGF>
F;928-930/Region: cell attachment (R-G-D) motif
F;151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status
F;167-226/Disulfide bonds: #status predicted
F;266,270/Disulfide bonds: interchain #status predicted
F;612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

  Query Match             79.5%;  Score 35;  DB 1;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 70;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYHL 7
          ||||:||
Db   1108 TAYRWHL 1114


RESULT 9
A70489
conserved hypothetical protein aq_2197 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 11-Jan-2000
C;Accession: A70489
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: A70489
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-162 <AQF>
A;Cross-references: GB:AE000777; NID:g2984377; PIDN:AAC07909.1; PID:g2984391; GB:AE
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_2197
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repe
F;11-43/Domain: tetratricopeptide repeat homology #status atypical <TT1>
F;44-77/Domain: tetratricopeptide repeat homology <TT2>
F;78-111/Domain: tetratricopeptide repeat homology <TT3>
F;114-147/Domain: tetratricopeptide repeat homology <TT4>

  Query Match             77.3%;  Score 34;  DB 2;  Length 162;
  Best Local Similarity   71.4%;  Pred. No. 14;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 AYRYHLL 8
          ||:||:|
Db     18 AYKYHML 24


RESULT 10
T17937
hypothetical protein a434L - Chlorella virus PBCV-1
C;Species: Chlorella virus PBCV-1
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17937
R;Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A;Reference number: Z18806
A;Accession: T17937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-72 <GRA>
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96802.1
A;Experimental source: specific host Chlorella strain NC64A
C;Genetics:
A;Note: a434L

  Query Match             75.0%;  Score 33;  DB 2;  Length 72;
  Best Local Similarity   71.4%;  Pred. No. 9.7;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 AYRYHLL 8
          |:|||:|
Db     55 AFRYHIL 61
```

```
RESULT 11
D72586
hypothetical protein APE1161 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C;Accession: D72586
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72586
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-173 <KAW>
A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80146.1; PID:d1043932; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1161
C;Superfamily: Aeropyrum pernix hypothetical protein APE1161

  Query Match             75.0%;  Score 33;  DB 2;  Length 173;
  Best Local Similarity   83.3%;  Pred. No. 24;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 YRYHLL 8
           |||| :|
Db     125 YRYHIL 130


RESULT 12
G70415
nucleotide sugar epimerase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C;Accession: G70415
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: G70415
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-321 <AQF>
A;Cross-references: GB:AE000735; NID:g2983749; PIDN:AAC07310.1; PID:g2983750; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: nse
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
F;3-316/Domain: UDPglucose 4-epimerase homology <UDP>

  Query Match             75.0%;  Score 33;  DB 2;  Length 321;
  Best Local Similarity   75.0%;  Pred. No. 46;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           ||| || |
Db     161 TAYTYHYL 168


RESULT 13
H81399
2-oxoglutarate-ferredoxin oxidoreductase (EC 1.2.7.-) alpha chain Cj0536 [similarity] -
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: H81399
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: H81399
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-374 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75172.1; PID:g69
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: oorA; Cj0536
C;Superfamily: Helicobacter pylori 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid fe
C;Keywords: oxidoreductase

  Query Match             75.0%;  Score 33;  DB 2;  Length 374;
  Best Local Similarity   71.4%;  Pred. No. 54;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           | ||||:
Db     220 TGYRYHV 226


RESULT 14
S27613
hypothetical protein 1 - Streptomyces aureofaciens (ATCC 10762) plasmid pOP2621 (fra
C;Species: Streptomyces aureofaciens
A;Variety: ATCC 10762
C;Date: 22-Nov-1993 #sequence_revision 23-Feb-1996 #text_change 15-Jun-2001
C;Accession: S27613
R;Pelletier, I.; Pfeifer, O.; Altenbuchner, J.; van Pee, K.H.
submitted to the EMBL Data Library, February 1992
A;Description: Molecular cloning and sequencing of a non haem bromoperoxidase from S
A;Reference number: S27613
A;Accession: S27613
A;Molecule type: DNA
A;Residues: 1-150 <PEL>
A;Cross-references: EMBL:M84990; NID:g150455; PIDN:AAB84314.1; PID:g150456
C;Genetics:
A;Genome: plasmid pOP2621

  Query Match             72.7%;  Score 32;  DB 2;  Length 150;
  Best Local Similarity   75.0%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           |||: |||
Db      64 TAYQLHLL 71


RESULT 15
AC2548
hypothetical protein alr7670 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC71
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AC2548
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2548
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-172 <KUR>
A;Cross-references: GB:AP003602; PIDN:BAB77313.1; PID:g17134755; GSPDB:GN00181
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr7670
A;Genome: plasmid

  Query Match             72.7%;  Score 32;  DB 2;  Length 172;
  Best Local Similarity   83.3%;  Pred. No. 38;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      2 AYRYHL 7
          |||||:
Db    136 AYRYHV 141

Search completed: May 22, 2003, 12:10:13
Job time : 16 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:57 ; Search time 7.5 Seconds
(without alignments)
44.241 Million cell updates/sec

Title: US-09-719-494-12
Perfect score: 44
Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	37	84.1	1172	1	TSP2_MOUSE	Q03350	mus musculu
2	37	84.1	1178	1	TSP2_CHICK	P35440	gallus gall
3	35	79.5	944	1	CHS2_NEUCR	P30589	neurospora
4	35	79.5	988	1	CHS1_EXODE	P30600	exophiala d
5	35	79.5	1043	1	CHS2_PARBR	Q92444	paracoccidi
6	35	79.5	1170	1	TSP1_BOVIN	Q28178	bos taurus
7	35	79.5	1170	1	TSP2_BOVIN	Q95116	bos taurus
8	35	79.5	1172	1	TSP2_HUMAN	P35442	homo sapien
9	35	79.5	1173	1	TSP1_XENLA	P35448	xenopus lae
10	32	72.7	150	1	YBP2_STRAU	P29714	streptomyce
11	32	72.7	234	1	OPAB_NEIGO	Q04874	neisseria g
12	32	72.7	234	1	OPAE_NEIGO	Q04878	neisseria g
13	32	72.7	234	1	OPAF_NEIGO	Q04879	neisseria g
14	32	72.7	260	1	OPR1_NEIMC	P10170	neisseria m
15	32	72.7	270	1	OMPC_NEIGO	P09888	neisseria g
16	32	72.7	517	1	TYR2_MOUSE	P29812	mus musculu
17	32	72.7	519	1	TYR2_HUMAN	P40126	homo sapien
18	32	72.7	522	1	TYR1_CARAU	P55028	carassius a
19	32	72.7	534	1	TYR1_AMBME	P55027	ambystoma m
20	32	72.7	535	1	TYR1_CHICK	O57405	gallus gall
21	32	72.7	537	1	TYR1_HUMAN	P17643	homo sapien
22	32	72.7	537	1	TYR1_MOUSE	P07147	mus musculu
23	32	72.7	712	1	Y352_HUMAN	O15060	homo sapien
24	32	72.7	3951	1	VGF1_IBVB	P27920	avian infec
25	31	70.5	72	1	VG83_BPML5	Q05300	mycobacteri
26	31	70.5	72	1	Y03I_BPT4	P39489	bacteriopha
27	31	70.5	120	1	PA23_BOTPI	P58464	bothrops pi
28	31	70.5	121	1	PA21_BOTPI	P58399	bothrops pi
29	31	70.5	121	1	PA22_BOTPI	P82287	bothrops pi
30	31	70.5	121	1	PA2H_BOTJR	Q90249	bothrops ja
31	31	70.5	151	1	TIM3_RABIT	O97590	oryctolagus
32	31	70.5	211	1	TIM3_BOVIN	P79121	bos taurus
33	31	70.5	211	1	TIM3_HORSE	Q9tul9	equus cabal
34	31	70.5	211	1	TIM3_HUMAN	P35625	homo sapien
35	31	70.5	211	1	TIM3_MOUSE	P39876	mus musculu
36	31	70.5	211	1	TIM3_RAT	P48032	rattus norv
37	31	70.5	231	1	VF16_VACCP	P29892	vaccinia vi
38	31	70.5	236	1	PHOU_XYLFA	Q9pbj9	xylella fas
39	31	70.5	269	1	DAPF_CHLMU	Q9pjw2	chlamydia m
40	31	70.5	280	1	TRUA_TREPA	O83802	treponema p
41	31	70.5	306	1	YBFH_BACSU	O31448	bacillus su
42	31	70.5	314	1	YDQC_SCHPO	O14204	schizosacch
43	31	70.5	349	1	YGA1_YEAST	P53199	s putative
44	31	70.5	421	1	YJ9E_YEAST	P47168	saccharomyc
45	31	70.5	566	1	PKNE_MYCTU	P72001	mycobacteri

ALIGNMENTS

RESULT 1
TSP2_MOUSE
ID TSP2_MOUSE STANDARD; PRT; 1172 AA.
AC Q03350;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT genome.";
RL J. Biol. Chem. 266:12821-12824(1991).
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC ----------

CC ----------
DR EMBL; L07803; AAA53064.1; -.
DR EMBL; M64866; AAA40432.1; -.
DR PIR; A42587; A42587.
DR PIR; A39851; A39851.
DR HSSP; P00740; 1EDM.
DR MGD; MGI:98738; Thbs2.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.

```
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 9.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19   1172       THROMBOSPONDIN 2.
FT   DOMAIN      19    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN     318    375       VWFC.
FT   DOMAIN     381    432       TSP TYPE-1 1.
FT   DOMAIN     437    493       TSP TYPE-1 2.
FT   DOMAIN     494    548       TSP TYPE-1 3.
FT   DOMAIN     549    589       EGF-LIKE 1.
FT   DOMAIN     590    647       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     648    692       EGF-LIKE 3.
FT   DOMAIN     725    760       TSP TYPE-3 1.
FT   DOMAIN     761    783       TSP TYPE-3 2.
FT   DOMAIN     784    819       TSP TYPE-3 3.
FT   DOMAIN     820    842       TSP TYPE-3 4.
FT   DOMAIN     843    880       TSP TYPE-3 5.
FT   DOMAIN     881    916       TSP TYPE-3 6.
FT   DOMAIN     917    952       TSP TYPE-3 7.
FT   DOMAIN     953   1172       C-TERMINAL.
FT   SITE       928    930       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID   266    266       INTERCHAIN (PROBABLE).
FT   DISULFID   270    270       INTERCHAIN (PROBABLE).
FT   DISULFID   553    564       BY SIMILARITY.
FT   DISULFID   558    574       BY SIMILARITY.
FT   DISULFID   577    588       BY SIMILARITY.
FT   DISULFID   594    610       BY SIMILARITY.
FT   DISULFID   601    619       BY SIMILARITY.
FT   DISULFID   622    646       BY SIMILARITY.
FT   DISULFID   652    665       BY SIMILARITY.
FT   DISULFID   659    678       BY SIMILARITY.
FT   DISULFID   680    691       BY SIMILARITY.
FT   CARBOHYD   151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   584    584       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1069   1069       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1172 AA;  129911 MW;  7CE8E4E8599822AB CRC64;

  Query Match             84.1%;  Score 37;  DB 1;  Length 1172;
  Best Local Similarity   75.0%;  Pred. No. 10;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           ||||:||:
Db    1108 TAYRWHLI 1115


RESULT 2
TSP2_CHICK
ID   TSP2_CHICK     STANDARD;      PRT;  1178 AA.
AC   P35440;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thrombospondin 2 precursor.
GN   THBS2 OR TSP2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91217026; PubMed=2022631;
RA   Lawler J., Duquette M., Ferro P.;
RT   "Cloning and sequencing of chicken thrombospondin.";
RL   J. Biol. Chem. 266:8039-8043(1991).
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ AND TYPE V COLLAGEN.
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M60853; AAA51437.1; -.
DR   PIR; A39804; A39804.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 8.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23   1178       THROMBOSPONDIN 2.
FT   DOMAIN       ?    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN     324    381       VWFC.
FT   DOMAIN     387    438       TSP TYPE-1 1.
FT   DOMAIN     441    499       TSP TYPE-1 2.
FT   DOMAIN     500    553       TSP TYPE-1 3.
FT   DOMAIN     555    595       EGF-LIKE 1.
FT   DOMAIN     596    653       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     654    698       EGF-LIKE 3.
FT   DOMAIN     731    766       TSP TYPE-3 1.
FT   DOMAIN     767    789       TSP TYPE-3 2.
FT   DOMAIN     790    825       TSP TYPE-3 3.
FT   DOMAIN     826    848       TSP TYPE-3 4.
FT   DOMAIN     849    886       TSP TYPE-3 5.
```

```
FT   DOMAIN       887    922       TSP TYPE-3 6.
FT   DOMAIN       923    958       TSP TYPE-3 7.
FT   DOMAIN       959   1178       C-TERMINAL.
FT   SITE         934    935       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     559    570       BY SIMILARITY.
FT   DISULFID     564    580       BY SIMILARITY.
FT   DISULFID     583    594       BY SIMILARITY.
FT   DISULFID     600    616       BY SIMILARITY.
FT   DISULFID     607    625       BY SIMILARITY.
FT   DISULFID     628    652       BY SIMILARITY.
FT   DISULFID     658    671       BY SIMILARITY.
FT   DISULFID     665    684       BY SIMILARITY.
FT   DISULFID     686    697       BY SIMILARITY.
FT   CARBOHYD     157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     317    317       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     322    322       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     463    463       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     590    590       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     716    716       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1075   1075       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1178 AA;  131816 MW;  F37E02F42C8717A2 CRC64;

  Query Match             84.1%;  Score 37;  DB 1;  Length 1178;
  Best Local Similarity   75.0%;  Pred. No. 10;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYHLL 8
          ||||:||:
Db   1114 TAYRWHLI 1121


RESULT 3
CHS2_NEUCR
ID   CHS2_NEUCR     STANDARD;      PRT;   944 AA.
AC   P30589;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Chitin synthase 2 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
DE   transferase 2).
GN   CHS-2.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95039879; PubMed=7952169;
RA   Din A.B., Yarden O.;
RT   "The Neurospora crassa chs-2 gene encodes a non-essential chitin
RT   synthase.";
RL   Microbiology 140:2189-2197(1994).
RN   [2]
RP   SEQUENCE OF 250-438 FROM N.A.
RX   MEDLINE=92115692; PubMed=1731323;
RA   Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
RA   Robbins P.W.;
RT   "Classification of fungal chitin synthases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
CC   -!- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC   -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
CC       beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-beta-D-
CC       glucosaminyl)}(N+1).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane.
CC   -!- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77782; CAA54816.1; -.
DR   EMBL; M82951; AAA33582.1; -.
DR   PIR; B45189; B45189.
DR   InterPro; IPR004834; Chitin_synth.
DR   Pfam; PF01644; Chitin_synth; 1.
DR   ProDom; PD002998; Chitin_synth; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW   Multigene family.
FT   TRANSMEM     597    617       POTENTIAL.
FT   TRANSMEM     634    654       POTENTIAL.
FT   TRANSMEM     669    689       POTENTIAL.
FT   TRANSMEM     713    733       POTENTIAL.
FT   TRANSMEM     873    893       POTENTIAL.
SQ   SEQUENCE   944 AA;  106816 MW;  F70052AEE083060D CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 944;
  Best Local Similarity   75.0%;  Pred. No. 20;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYHLL 8
          :|||||:|
Db    468 SAYRYHAL 475


RESULT 4
CHS1_EXODE
ID   CHS1_EXODE     STANDARD;      PRT;   988 AA.
AC   P30600; O74681;
DT   01-APR-1993 (Rel. 25, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Chitin synthase 1 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
DE   transferase 1) (Class-II chitin synthase 1).
GN   CHS1.
OS   Exophiala dermatitidis (Wangiella dermatitidis).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes;
OC   Chaetothyriales; Herpotrichiellaceae; mitosporic Herpotrichiellaceae;
OC   Exophiala.
OX   NCBI_TaxID=5970;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=8656;
RA   Zheng L., Mendoza A.L., Wang Z., Szaniszlo P.J.;
RT   "Characterization of WdCHS1, a class II chitin synthase gene, and
RT   multiple wdchs mutants of Wangiella dermatitidis.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 265-468 FROM N.A.
RX   MEDLINE=92115692; PubMed=1731323;
RA   Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
RA   Robbins P.W.;
RT   "Classification of fungal chitin synthases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
CC   -!- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC   -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
CC       beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-beta-D-
CC       glucosaminyl)}(N+1).
CC   -!- SUBCELLULAR LOCATION: Plasma membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC       II.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF054503; AAC36064.1; -.
DR   EMBL; M81905; AAA30334.1; -.
```

```
DR   InterPro; IPR004834; Chitin_synth.
DR   Pfam; PF01644; Chitin_synth; 1.
DR   ProDom; PD002998; Chitin_synth; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW   Multigene family.
FT   CONFLICT    265    266       CV -> KL (IN REF. 2).
FT   CONFLICT    468    468       I -> L (IN REF. 2).
SQ   SEQUENCE   988 AA;  111220 MW;  5940234B344F88A4 CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 988;
  Best Local Similarity   75.0%;  Pred. No. 21;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           :|||||  |
Db     491 SAYRYHAL 498


RESULT 5
CHS2_PARBR
ID   CHS2_PARBR     STANDARD;      PRT;  1043 AA.
AC   Q92444;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Chitin synthase 2 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
DE   transferase 2) (Class-II chitin synthase 2).
GN   CHS2.
OS   Paracoccidioides brasiliensis.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Onygenales; mitosporic Onygenales; Paracoccidioides.
OX   NCBI_TaxID=121759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IVIC Pb73 / ATCC 32071;
RX   MEDLINE=98144800; PubMed=9483806;
RA   Nino-Vega G.A., Buurman E.T., Gooday G.W., San-Blas G., Gow N.A.R.;
RT   "Molecular cloning and sequencing of a chitin synthase gene (CHS2) of
RT   Paracoccidioides brasiliensis.";
RL   Yeast 14:181-187(1998).
CC   -!- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC   -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
CC       beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-beta-D-
CC       glucosaminyl)}(N+1).
CC   -!- SUBCELLULAR LOCATION: Plasma membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
CC       II.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y09231; CAA70433.1; -.
DR   InterPro; IPR004834; Chitin_synth.
DR   Pfam; PF01644; Chitin_synth; 1.
DR   ProDom; PD002998; Chitin_synth; 1.
KW   Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW   Multigene family.
FT   TRANSMEM    663    683       POTENTIAL.
FT   TRANSMEM    703    723       POTENTIAL.
FT   TRANSMEM    738    758       POTENTIAL.
FT   TRANSMEM    780    800       POTENTIAL.
FT   TRANSMEM    907    927       POTENTIAL.
FT   TRANSMEM    931    951       POTENTIAL.
SQ   SEQUENCE   1043 AA;  116587 MW;  F63410221B325B73 CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 1043;
  Best Local Similarity   75.0%;  Pred. No. 23;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           :|||||  |
Db     538 SAYRYHAL 545


RESULT 6
TSP1_BOVIN
ID   TSP1_BOVIN     STANDARD;      PRT;  1170 AA.
AC   Q28178; Q28179;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Thrombospondin 1 precursor.
GN   THBS1 OR TSP1 OR TSP-1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Tooth;
RX   MEDLINE=98173773; PubMed=9507054;
RA   Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
RA   Inoue H.;
RT   "cDNA cloning of bovine thrombospondin 1 and its expression in
RT   odontoblasts and predentin.";
RL   Biochim. Biophys. Acta 1382:17-22(1998).
RN   [2]
RP   SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC   TISSUE=Aortic endothelium;
RA   Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC       V/BETA-3 AND ALPHA-IIB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS
CC       AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- TISSUE SPECIFICITY: ODONTOBLASTS.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005287; BAA21115.1; -.
DR   EMBL; X87618; CAA60950.1; -.
DR   EMBL; X87619; CAA60951.1; -.
DR   GlycoSuiteDB; Q28178; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 8.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00209; TSP1; 3.
```

```
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
DR   PROSITE; PS50092; TSP1; 3.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   CHAIN        19   1170       THROMBOSPONDIN 1.
FT   DOMAIN       19    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      316    373       VWFC.
FT   DOMAIN      379    430       TSP TYPE-1 1.
FT   DOMAIN      435    491       TSP TYPE-1 2.
FT   DOMAIN      492    548       TSP TYPE-1 3.
FT   DOMAIN      549    587       EGF-LIKE 1.
FT   DOMAIN      588    645       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      646    690       EGF-LIKE 3.
FT   DOMAIN      723    758       TSP TYPE-3 1.
FT   DOMAIN      759    781       TSP TYPE-3 2.
FT   DOMAIN      782    817       TSP TYPE-3 3.
FT   DOMAIN      818    840       TSP TYPE-3 4.
FT   DOMAIN      841    878       TSP TYPE-3 5.
FT   DOMAIN      879    914       TSP TYPE-3 6.
FT   DOMAIN      915    950       TSP TYPE-3 7.
FT   DOMAIN      951   1170       C-TERMINAL.
FT   SITE        926    928       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    270    270       INTERCHAIN (PROBABLE).
FT   DISULFID    274    274       INTERCHAIN (PROBABLE).
FT   DISULFID    551    562       BY SIMILARITY.
FT   DISULFID    556    572       BY SIMILARITY.
FT   DISULFID    575    586       BY SIMILARITY.
FT   DISULFID    592    608       BY SIMILARITY.
FT   DISULFID    599    617       BY SIMILARITY.
FT   DISULFID    620    644       BY SIMILARITY.
FT   DISULFID    650    663       BY SIMILARITY.
FT   DISULFID    657    676       BY SIMILARITY.
FT   DISULFID    678    689       BY SIMILARITY.
FT   CARBOHYD    248    248       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1067   1067       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1085   1085       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    805    805       S -> G (IN REF. 2).
SQ   SEQUENCE   1170 AA;  129533 MW;  0DD6ADF3E5FA031A CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 1170;
  Best Local Similarity   85.7%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
             ||||:||
Db    1106 TAYRWHL 1112


RESULT 7
TSP2_BOVIN
ID   TSP2_BOVIN     STANDARD;      PRT;  1170 AA.
AC   Q95116; Q28180;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE   (CISP).
GN   THBS2 OR TSP2 OR TSP-2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Danik M., Chinn A., Lafeuillade M., Keramidas M., Aguesse-Germon S.,
RA   Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-522 FROM N.A.
RX   MEDLINE=96331130; PubMed=8698834;
RA   Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA   Feige J.J.;
RT   "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT   secreted protein/thrombospondin-2 expression by adrenocorticotropic
RT   hormone in adrenocortical cells.";
RL   J. Cell. Physiol. 167:164-172(1996).
RN   [3]
RP   SEQUENCE OF 318-831 FROM N.A.
RC   TISSUE=Aortic endothelium;
RA   Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT   "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT   TGF-beta.";
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ AND TYPE V COLLAGEN.
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X96540; CAA65385.1; -.
DR   EMBL; X87620; CAA60952.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 9.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1170       THROMBOSPONDIN 2.
FT   DOMAIN       19    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      318    375       VWFC.
FT   DOMAIN      379    430       TSP TYPE-1 1.
FT   DOMAIN      435    491       TSP TYPE-1 2.
FT   DOMAIN      492    546       TSP TYPE-1 3.
FT   DOMAIN      547    587       EGF-LIKE 1.
FT   DOMAIN      588    645       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      646    690       EGF-LIKE 3.
FT   DOMAIN      723    758       TSP TYPE-3 1.
FT   DOMAIN      759    781       TSP TYPE-3 2.
```

```
FT   DOMAIN      782    817       TSP TYPE-3 3.
FT   DOMAIN      818    840       TSP TYPE-3 4.
FT   DOMAIN      841    878       TSP TYPE-3 5.
FT   DOMAIN      879    914       TSP TYPE-3 6.
FT   DOMAIN      915    950       TSP TYPE-3 7.
FT   DOMAIN      951   1170       C-TERMINAL.
FT   SITE        926    928       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    266    266       INTERCHAIN (PROBABLE).
FT   DISULFID    270    270       INTERCHAIN (PROBABLE).
FT   DISULFID    551    562       BY SIMILARITY.
FT   DISULFID    556    572       BY SIMILARITY.
FT   DISULFID    575    586       BY SIMILARITY.
FT   DISULFID    592    608       BY SIMILARITY.
FT   DISULFID    599    617       BY SIMILARITY.
FT   DISULFID    620    644       BY SIMILARITY.
FT   DISULFID    650    663       BY SIMILARITY.
FT   DISULFID    657    676       BY SIMILARITY.
FT   DISULFID    678    689       BY SIMILARITY.
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    455    455       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    582    582       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    936    936       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1067   1067       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    535    535       A -> V (IN REF. 3).
FT   CONFLICT    748    748       S -> T (IN REF. 3).
SQ   SEQUENCE   1170 AA;  129862 MW;  9CF1FBF55B89A051 CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 1170;
  Best Local Similarity   85.7%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYHL 7
            ||||:||
Db   1106 TAYRWHL 1112


RESULT 8
TSP2_HUMAN
ID   TSP2_HUMAN     STANDARD;      PRT;  1172 AA.
AC   P35442;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thrombospondin 2 precursor.
GN   THBS2 OR TSP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94010892; PubMed=8406456;
RA   Labell T.L., Byers P.H.;
RT   "Sequence and characterization of the complete human thrombospondin 2
RT   cDNA: potential regulatory role for the 3' untranslated region.";
RL   Genomics 17:225-229(1993).
RN   [2]
RP   SEQUENCE OF 560-1172 FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=92217961; PubMed=1559694;
RA   Labell T.L., McGookey Milewicz D.J., Disteche C.M., Byers P.H.;
RT   "Thrombospondin II: partial cDNA sequence, chromosome location, and
RT   expression of a second member of the thrombospondin gene family in
RT   humans.";
RL   Genomics 12:421-429(1992).
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ AND TYPE V COLLAGEN.
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L12350; AAA03703.1; -.
DR   EMBL; M81339; -; NOT_ANNOTATED_CDS.
DR   PIR; A42173; A42173.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:11786; THBS2.
DR   MIM; 188061; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 9.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1172       THROMBOSPONDIN 2.
FT   DOMAIN       19    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      318    375       VWFC.
FT   DOMAIN      381    432       TSP TYPE-1 1.
FT   DOMAIN      437    493       TSP TYPE-1 2.
FT   DOMAIN      494    548       TSP TYPE-1 3.
FT   DOMAIN      549    589       EGF-LIKE 1.
FT   DOMAIN      590    647       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      648    692       EGF-LIKE 3.
FT   DOMAIN      725    760       TSP TYPE-3 1.
FT   DOMAIN      761    783       TSP TYPE-3 2.
FT   DOMAIN      784    819       TSP TYPE-3 3.
FT   DOMAIN      820    842       TSP TYPE-3 4.
FT   DOMAIN      843    880       TSP TYPE-3 5.
FT   DOMAIN      881    916       TSP TYPE-3 6.
FT   DOMAIN      917    952       TSP TYPE-3 7.
FT   DOMAIN      953   1172       C-TERMINAL.
FT   SITE        928    930       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    266    266       INTERCHAIN (PROBABLE).
FT   DISULFID    270    270       INTERCHAIN (PROBABLE).
FT   DISULFID    553    564       BY SIMILARITY.
FT   DISULFID    558    574       BY SIMILARITY.
FT   DISULFID    577    588       BY SIMILARITY.
FT   DISULFID    594    610       BY SIMILARITY.
FT   DISULFID    601    619       BY SIMILARITY.
FT   DISULFID    622    646       BY SIMILARITY.
FT   DISULFID    652    665       BY SIMILARITY.
FT   DISULFID    659    678       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    584    584       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1069   1069       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1172 AA;  129955 MW;  2AC7BB230E44C6F5 CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 1172;
  Best Local Similarity   85.7%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1108 TAYRWHL 1114


RESULT 9
TSP1_XENLA
ID   TSP1_XENLA     STANDARD;      PRT;  1173 AA.
AC   P35448;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thrombospondin 1 precursor.
GN   THBS1 OR TSP1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
RL   Submitted (XXX-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC       V/BETA-3 AND ALPHA-IIB/BETA-3 (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04278; -; NOT_ANNOTATED_CDS.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 8.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23   1173       THROMBOSPONDIN 1.
FT   DOMAIN       23    235       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      319    376       VWFC.
FT   DOMAIN      382    433       TSP TYPE-1 1.
FT   DOMAIN      438    494       TSP TYPE-1 2.
FT   DOMAIN      495    546       TSP TYPE-1 3.
FT   DOMAIN      550    590       EGF-LIKE 1.
FT   DOMAIN      591    648       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      649    693       EGF-LIKE 3.
FT   DOMAIN      726    761       TSP TYPE-3 1.
FT   DOMAIN      762    784       TSP TYPE-3 2.
FT   DOMAIN      785    820       TSP TYPE-3 3.
FT   DOMAIN      821    843       TSP TYPE-3 4.
FT   DOMAIN      844    881       TSP TYPE-3 5.
FT   DOMAIN      882    917       TSP TYPE-3 6.
FT   DOMAIN      918    953       TSP TYPE-3 7.
FT   DOMAIN      954   1173       C-TERMINAL.
FT   SITE        929    931       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    554    565       BY SIMILARITY.
FT   DISULFID    559    575       BY SIMILARITY.
FT   DISULFID    578    589       BY SIMILARITY.
FT   DISULFID    595    611       BY SIMILARITY.
FT   DISULFID    602    620       BY SIMILARITY.
FT   DISULFID    623    647       BY SIMILARITY.
FT   DISULFID    653    666       BY SIMILARITY.
FT   DISULFID    660    679       BY SIMILARITY.
FT   DISULFID    681    692       BY SIMILARITY.
FT   CARBOHYD    155    155       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    158    158       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    363    363       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    705    705       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    711    711       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1070   1070       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1173 AA;  130019 MW;  A9F036D6516C0F24 CRC64;

  Query Match             79.5%;  Score 35;  DB 1;  Length 1173;
  Best Local Similarity   85.7%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 TAYRYHL 7
           ||||:||
Db    1109 TAYRWHL 1115


RESULT 10
YBP2_STRAU
ID   YBP2_STRAU     STANDARD;      PRT;   150 AA.
AC   P29714;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein in BPOA2 5'region (Fragment).
OS   Streptomyces aureofaciens.
OG   Plasmid pOP2621.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1894;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10762 / CCM 3239;
RX   MEDLINE=92407485; PubMed=1527491;
RA   Pfeifer O., Pelletier I., Altenbuchner J., van Pee K.-H.;
RT   "Molecular cloning and sequencing of a non-haem bromoperoxidase gene
RT   from Streptomyces aureofaciens ATCC 10762.";
RL   J. Gen. Microbiol. 138:1123-1131(1992).
CC   -!- SIMILARITY: BELONGS TO THE DPS FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M84990; AAB84314.1; -.
DR   PIR; S27613; S27613.
DR   InterPro; IPR002177; DPS.
DR   Pfam; PF02047; DPS; 1.
DR   ProDom; PD149803; DPS; 1.
DR   PROSITE; PS00818; DPS_1; 1.
DR   PROSITE; PS00819; DPS_2; 1.
KW   Hypothetical protein; Plasmid.
FT   NON_TER     150    150
SQ   SEQUENCE   150 AA;  16648 MW;  45EF8CD9573E875C CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 150;
  Best Local Similarity   75.0%;  Pred. No. 10;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYHLL 8
           |||: |||
Db      64 TAYQLHLL 71


RESULT 11
OPAB_NEIGO
ID   OPAB_NEIGO     STANDARD;      PRT;   234 AA.
AC   Q04874;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Opacity protein OPA51 precursor (Fragment).
GN   OPAB.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS11 / F3;
RX   MEDLINE=93178439; PubMed=8440254;
RA   Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT   "Variable opacity (Opa) outer membrane proteins account for the cell
RT   tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT   epithelial cells.";
RL   EMBO J. 12:641-650(1993).
CC   -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC       PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC       VARIATION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z18928; CAA79361.1; -.
DR   PIR; S28628; S28628.
DR   PIR; S36329; S36329.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      1       POTENTIAL.
FT   CHAIN         2    >234      OPACITY PROTEIN OPA51.
FT   NON_TER     234    234
SQ   SEQUENCE   234 AA;  26772 MW;  9FFE5B5DABBA96CA CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 234;
  Best Local Similarity   83.3%;  Pred. No. 17;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYH 6
           | ||||
Db     206 TGYRYH 211


RESULT 12
OPAE_NEIGO
ID   OPAE_NEIGO     STANDARD;      PRT;   234 AA.
AC   Q04878;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Opacity protein OPA55 precursor (Fragment).
GN   OPAE.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS11 / F3;
RX   MEDLINE=93178439; PubMed=8440254;
RA   Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT   "Variable opacity (Opa) outer membrane proteins account for the cell
RT   tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT   epithelial cells.";
RL   EMBO J. 12:641-650(1993).
CC   -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC       PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC       VARIATION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z18933; CAA79366.1; -.
DR   PIR; S28632; S28632.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      1       POTENTIAL.
FT   CHAIN         2    >234      OPACITY PROTEIN OPA55.
FT   NON_TER     234    234
SQ   SEQUENCE   234 AA;  26881 MW;  8EEB30B3A774C766 CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 234;
  Best Local Similarity   83.3%;  Pred. No. 17;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 TAYRYH 6
           | ||||
Db     206 TGYRYH 211


RESULT 13
OPAF_NEIGO
ID   OPAF_NEIGO     STANDARD;      PRT;   234 AA.
AC   Q04879;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Opacity protein OPA56 precursor (Fragment).
```

```
GN   OPAF.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS11 / F3;
RX   MEDLINE=93178439; PubMed=8440254;
RA   Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT   "Variable opacity (Opa) outer membrane proteins account for the cell
RT   tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT   epithelial cells.";
RL   EMBO J. 12:641-650(1993).
CC   -!- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC       PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC       VARIATION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z18934; CAA79367.1; -.
DR   PIR; S28620; S28620.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1      1       POTENTIAL.
FT   CHAIN         2    >234      OPACITY PROTEIN OPA56.
FT   NON_TER     234    234
SQ   SEQUENCE   234 AA;  26868 MW;  5175C6606839EFFB CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 234;
  Best Local Similarity   83.3%;  Pred. No. 17;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          | ||||
Db    206 TGYRYH 211


RESULT 14
OPR1_NEIMC
ID   OPR1_NEIMC     STANDARD;      PRT;   260 AA.
AC   P10170;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opacity-related protein POPM1.
GN   OPR.
OS   Neisseria meningitidis (serogroup C).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=135720;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C1938 / Serogroup C;
RX   MEDLINE=88260884; PubMed=2455211;
RA   Stern A., Meyer T.F.;
RT   "Common mechanism controlling phase and antigenic variation in
RT   pathogenic neisseriae.";
RL   Mol. Microbiol. 1:5-12(1987).
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: STRONG TO THE OPACITY-RELATED PROTEIN POPM3 AND
CC       REGIONS OF HOMOLOGY WITH N.GONORRHOEAE (STRAIN MS11) OPA GENE
CC       PRODUCTS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06445; CAA29748.1; ALT_SEQ.
DR   PIR; S08514; S08514.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane.
SQ   SEQUENCE   260 AA;  28936 MW;  EB47A2843B3F037B CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 260;
  Best Local Similarity   83.3%;  Pred. No. 19;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          | ||||
Db    232 TGYRYH 237


RESULT 15
OMPC_NEIGO
ID   OMPC_NEIGO     STANDARD;      PRT;   270 AA.
AC   P09888;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Outer membrane protein P.IIC precursor (Protein IIC).
GN   PIIC.
OS   Neisseria gonorrhoeae.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=485;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JS3;
RX   MEDLINE=89096501; PubMed=3145386;
RA   van der Ley P.;
RT   "Three copies of a single protein II-encoding sequence in the genome
RT   of Neisseria gonorrhoeae JS3: evidence for gene conversion and gene
RT   duplication.";
RL   Mol. Microbiol. 2:797-806(1988).
CC   -!- FUNCTION: THIS PROTEIN SERVES AS A PORIN.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12625; CAA31144.1; -.
DR   PIR; S03095; KONH2C.
DR   InterPro; IPR003394; Porin_opacity.
DR   Pfam; PF02462; Opacity; 1.
KW   Outer membrane; Porin; Transmembrane; Antigen; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26    270       OUTER MEMBRANE PROTEIN P.IIC.
FT   TRANSMEM     36     44       POTENTIAL.
FT   TRANSMEM     77     85       POTENTIAL.
FT   TRANSMEM     90     96       POTENTIAL.
FT   TRANSMEM    143    157       POTENTIAL.
FT   TRANSMEM    163    173       POTENTIAL.
FT   TRANSMEM    222    234       POTENTIAL.
FT   TRANSMEM    238    246       POTENTIAL.
FT   TRANSMEM    262    270       POTENTIAL.
SQ   SEQUENCE   270 AA;  30269 MW;  F6B448373830A50D CRC64;

  Query Match             72.7%;  Score 32;  DB 1;  Length 270;
```

Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 TAYRYH 6
           | ||||
Db     242 TGYRYH 247
```

Search completed: May 22, 2003, 12:08:40
Job time : 8.5 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:04:33 ; Search time 25 Seconds
(without alignments)
65.935 Million cell updates/sec

Title: US-09-719-494-12
Perfect score: 44
Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	81.8	145	16	Q9RT39	Q9rt39 deinococcus
2	36	81.8	262	10	Q8W2Q2	Q8w2q2 oryza sativ
3	36	81.8	569	10	Q8W2Q4	Q8w2q4 oryza sativ
4	36	81.8	1199	6	Q28205	Q28205 bos taurus
5	36	81.8	1372	2	Q54151	Q54151 shigella fl
6	36	81.8	1373	2	Q9AL58	Q9al58 shigella fl
7	36	81.8	1714	10	Q94HT9	Q94ht9 oryza sativ
8	36	81.8	1731	10	Q8SB47	Q8sb47 oryza sativ
9	36	81.8	1773	10	Q94HV6	Q94hv6 oryza sativ
10	36	81.8	1777	10	Q94HP9	Q94hp9 oryza sativ
11	36	81.8	1785	10	Q8S789	Q8s789 oryza sativ
12	36	81.8	1819	10	Q8SAZ4	Q8saz4 oryza sativ
13	36	81.8	2162	10	Q9AYC2	Q9ayc2 oryza sativ
14	35	79.5	211	3	Q9HF73	Q9hf73 ashbya goss
15	35	79.5	323	16	Q8UAD4	Q8uad4 agrobacteri
16	35	79.5	967	3	Q9Y7H9	Q9y7h9 phaeosphaer
17	35	79.5	1064	3	Q9C400	Q9c400 coccidioide
18	34	77.3	162	16	O67941	O67941 aquifex aeo
19	34	77.3	361	10	Q94IU8	Q94iu8 arabidopsis
20	33	75.0	72	12	Q98485	Q98485 paramecium
21	33	75.0	112	5	Q9U2Z9	Q9u2z9 caenorhabdi
22	33	75.0	173	17	Q9YCV1	Q9ycv1 aeropyrum p
23	33	75.0	321	16	O67354	O67354 aquifex aeo
24	33	75.0	374	16	Q9PHX8	Q9phx8 campylobact
25	33	75.0	581	5	Q95TG6	Q95tg6 drosophila
26	33	75.0	1766	5	Q9V591	Q9v591 drosophila
27	32	72.7	71	4	Q9P1E7	Q9p1e7 homo sapien
28	32	72.7	119	10	Q9FTY3	Q9fty3 oryza sativ
29	32	72.7	148	2	O31167	O31167 streptomyce
30	32	72.7	172	16	Q8ZS42	Q8zs42 anabaena sp
31	32	72.7	197	5	Q93103	Q93103 aedes aegyp
32	32	72.7	214	2	Q51086	Q51086 neisseria m
33	32	72.7	230	2	Q9R9A7	Q9r9a7 neisseria m
34	32	72.7	232	2	Q9R9A9	Q9r9a9 neisseria m
35	32	72.7	232	2	Q9R9A8	Q9r9a8 neisseria m
36	32	72.7	232	2	Q9K4T9	Q9k4t9 neisseria l
37	32	72.7	232	2	Q9K4T7	Q9k4t7 neisseria l
38	32	72.7	232	2	Q9K4T3	Q9k4t3 neisseria s
39	32	72.7	232	2	Q9R3P5	Q9r3p5 neisseria m
40	32	72.7	233	2	Q9K4T5	Q9k4t5 neisseria l
41	32	72.7	234	2	O07280	O07280 neisseria m
42	32	72.7	234	2	Q9R7I8	Q9r718 neisseria m
43	32	72.7	234	2	O30759	O30759 neisseria m
44	32	72.7	234	2	O07287	O07287 neisseria m
45	32	72.7	234	2	Q9R7I9	Q9r719 neisseria m

ALIGNMENTS

RESULT 1
Q9RT39
ID Q9RT39 PRELIMINARY; PRT; 145 AA.
AC Q9RT39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSPOSASE.
GN DR1927.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002031; AAF11477.1; -.
DR TIGR; DR1927; -.
DR InterPro; IPR002686; Transposase_17.
DR Pfam; PF01797; Transposase_17; 1.
DR ProDom; PD003831; Transposase_17; 1.
KW Complete proteome.
SQ SEQUENCE 145 AA; 16950 MW; 2CD5098D1012CFF9 CRC64;

Query Match 81.8%; Score 36; DB 16; Length 145;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 TAYRYHLL 8
          | ||||:|
Db     22 TKYRYHVL 29


RESULT 2
Q8W2Q2
ID   Q8W2Q2      PRELIMINARY;      PRT;   262 AA.
AC   Q8W2Q2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative polyprotein.
GN   OSJNBA0028C16.20.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RT   "Rice Genomic Sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC098565; AAL69438.1; -.
KW   Polyprotein.
SQ   SEQUENCE   262 AA;  27461 MW;  C7BA920CF2486429 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 262;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     36 TAYRYH 41


RESULT 3
Q8W2Q4
ID   Q8W2Q4      PRELIMINARY;      PRT;   569 AA.
AC   Q8W2Q4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative polyprotein.
GN   OSJNBA0028C16.18.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RT   "Rice Genomic Sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC098565; AAL69436.1; -.
KW   Polyprotein.
SQ   SEQUENCE   569 AA;  60552 MW;  64ED54E8612BE9D7 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 569;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     99 TAYRYH 104


RESULT 4
Q28205
ID   Q28205      PRELIMINARY;      PRT;  1199 AA.
AC   Q28205;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Cofactor D.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96319731; PubMed=8706133;
RA   Tian G., Huang Y., Rommelaere H., Vandekerckhove J., Ampe C.,
RA   Cowan N.J.;
RT   "Pathway leading to correctly folded beta-tubulin.";
RL   Cell 86:287-296(1996).
DR   EMBL; U61233; AAB17537.1; -.
SQ   SEQUENCE   1199 AA;  133014 MW;  2FCF9CCA972599BA CRC64;

  Query Match             81.8%;  Score 36;  DB 6;  Length 1199;
  Best Local Similarity   85.7%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 AYRYHLL 8
          |||||:|
Db    985 AYRYHVL 991


RESULT 5
Q54151
ID   Q54151      PRELIMINARY;      PRT;  1372 AA.
AC   Q54151;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   HEMAGGLUTININ.
GN   SHE OR PIC.
OS   Shigella flexneri, and
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=623, 562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=SEROTYPE 2A, AND M4243A;
RA   Noriega F.R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=042;
RA   Henderson I.R., Czeczulin J.R., Nataro J.P.;
RT   "A second autotransporter protease from enteroaggregative Escherichia
RT   coli.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U35656; AAB58244.1; -.
DR   EMBL; AF097644; AAD23953.1; -.
DR   MEROPS; S06.005; -.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR004899; Pertact_sup.
DR   Pfam; PF02395; IGA1; 1.
DR   Pfam; PF03212; Pertactin; 1.
KW   Protease; Signal.
SQ   SEQUENCE   1372 AA;  146449 MW;  68CE71254EAC26EA CRC64;

  Query Match             81.8%;  Score 36;  DB 2;  Length 1372;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYHL 7
          | |||||
Db   1227 TGYRYHL 1233
```

```
RESULT 6
Q9AL58
ID   Q9AL58      PRELIMINARY;      PRT;  1373 AA.
AC   Q9AL58;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Pic.
GN   PIC.
OS   Shigella flexneri 2a.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=42897;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YSH6000T;
RX   MEDLINE=20231776; PubMed=10768931;
RA   Al-Hasani K., Henderson I.R., Sakellaris H., Rajakumar K., Grant T.,
RA   Nataro J.P., Robins-Browne R., Adler B.;
RT   "The sigA gene which is borne on the she pathogenicity island of
RT   Shigella flexneri 2a encodes an exported cytopathic protease involved
RT   in intestinal fluid accumulation.";
RL   Infect. Immun. 68:2457-2463(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YSH6000T;
RX   MEDLINE=21105956; PubMed=11162180;
RA   Al-Hasani K., Rajakumar K., Bulach D., Robins-Browne R., Adler B.,
RA   Sakellaris H.;
RT   "Genetic organization of the she pathogenicity island in Shigella
RT   flexneri 2a.";
RL   Microb. Pathog. 30:1-8(2001).
DR   EMBL; AF200692; AAK00464.1; -.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR004899; Pertact_sup.
DR   Pfam; PF02395; IGA1; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR00921; IGASERPTASE.
SQ   SEQUENCE   1373 AA;  146548 MW;  45C016CDAA577763 CRC64;

  Query Match             81.8%;  Score 36;  DB 2;  Length 1373;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 TAYRYHL 7
          | |||||
Db   1228 TGYRYHL 1234


RESULT 7
Q94HT9
ID   Q94HT9      PRELIMINARY;      PRT;  1714 AA.
AC   Q94HT9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative retroelement.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,
RA   Nascimento L., Preston R., Kirchoff K., King L., Vil M.D., Baker J.,
RA   Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,
RA   Cunnius D.M., Balija V., Shah R., Bahret A., O'Shaughnessy A.,
RA   Palmer L., Yang C., Dedhia N., McCombie W.R.;
RT   "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT   OSJNBb0036B06, from Chromosome 10, complete sequence.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC051624; AAK92560.1; -.
DR   InterPro; IPR001995; Aspprotease_rtrv.
DR   InterPro; IPR001969; Aspprotease_site.
DR   InterPro; IPR005162; Retrotrans_gag.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF03732; Retrotrans_gag; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; rvp; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   1714 AA;  194668 MW;  168C0139B37AF1E0 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1714;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     36 TAYRYH 41


RESULT 8
Q8SB47
ID   Q8SB47      PRELIMINARY;      PRT;  1731 AA.
AC   Q8SB47;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative polyprotein.
GN   OSJNBB0091009.12.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RT   "Rice Genomic Sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC091732; AAL77161.1; -.
KW   Polyprotein.
SQ   SEQUENCE   1731 AA;  195964 MW;  F0996E6D1D029FB0 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1731;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db    103 TAYRYH 108


RESULT 9
Q94HV6
ID   Q94HV6      PRELIMINARY;      PRT;  1773 AA.
AC   Q94HV6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative retroelement.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,
RA   Nascimento L., Preston R., Kirchoff K., King L., Vil M.D., Baker J.,
RA   Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,
RA   Cunnius D.M., Balija V., Shah R., Bahret A., O'Shaughnessy A.,
RA   Palmer L., Yang C., Dedhia N., McCombie W.R.;
RT   "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT   OSJNBb0036B06, from Chromosome 10, complete sequence.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC051624; AAK92543.1; -.
DR   InterPro; IPR001969; Aspprotease_site.
DR   InterPro; IPR005162; Retrotrans_gag.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF03732; Retrotrans_gag; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   1773 AA;  201241 MW;  958C0F0BF77A84D3 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1773;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     99 TAYRYH 104


RESULT 10
Q94HP9
ID   Q94HP9      PRELIMINARY;      PRT;  1777 AA.
AC   Q94HP9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative retroelement.
GN   OSJNBA0004A10.7.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Nascimento L., Spiegel L., de la Bastide M., Kirchoff K., King L.,
RA   Preston R., Vil M.D., Baker J., Zutavern T., Santos L., Bell M.,
RA   Miller B., Kuit K., Rodriguez S., Cunnius D.M., Balija V., Shah R.,
RA   Bahret A., Palmer L., Yang C., O'Shaughnessy A., Dedhia N.,
RA   McCombie W.R.;
RT   "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT   OSJNBa0065C16, from chromosome 10, complete sequence.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   McCombie W.R., de la Bastide M., Spiegel L., Nascimento L., Balija V.,
RA   Zutavern T., Bell M., Preston R., Kirchoff K., Kuit K., Baker J.,
RA   Santos L., Miller B., Cunnius D.M., Katzenberger F., Muller S.,
RA   Shah R., King L., Yang C., Dike S., O'Shaughnessy A., Palmer L.,
RA   Dedhia N.;
RT   "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT   OSJNBa0004A10, from chromosome 10, complete sequence.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC074354; AAK92588.1; -.
DR   EMBL; AC098682; AAM01103.1; -.
DR   InterPro; IPR001995; Aspprotease_rtrv.
DR   InterPro; IPR001969; Aspprotease_site.
DR   InterPro; IPR005162; Retrotrans_gag.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF03732; Retrotrans_gag; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; rvp; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   1777 AA;  201593 MW;  CAE2E6EBC53C08E9 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1777;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     99 TAYRYH 104


RESULT 11
Q8S789
ID   Q8S789      PRELIMINARY;      PRT;  1785 AA.
AC   Q8S789;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative retroelement.
GN   OSJNBA0093I09.7.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Spiegel L., de la Bastide M., Kirchoff K., Preston R., Kuit K.,
RA   Nascimento L., Baker J., Vil M.D., Zutavern T., Santos L., Miller B.,
RA   Cunnius D.M., Balija V., Shah R., King L., Bell M., Yang C., Dike S.,
RA   Palmer L., O'Shaughnessy A., Dedhia N., McCombie W.R.;
RT   "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT   OSJNBa0093I09, complete sequence.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC090486; AAM08795.1; -.
SQ   SEQUENCE   1785 AA;  202400 MW;  D38E671CAFB46293 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1785;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     99 TAYRYH 104


RESULT 12
Q8SAZ4
ID   Q8SAZ4      PRELIMINARY;      PRT;  1819 AA.
AC   Q8SAZ4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative polyprotein.
GN   OSJNBA0029P16.17.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J., Wilson R., Minx P., Du H.;
RT   "Rice Genomic Sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC093568; AAL78107.1; -.
KW   Polyprotein.
SQ   SEQUENCE   1819 AA;  205815 MW;  6720371C4CFB0BA7 CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 1819;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db     99 TAYRYH 104


RESULT 13
Q9AYC2
ID   Q9AYC2      PRELIMINARY;      PRT;  2162 AA.
AC   Q9AYC2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Polyprotein.
GN   OSJNBA0094H10.13.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Spiegel L.A., Nascimento L.U., de la Bastide M., Kirchoff K.A.,
RA   King L., Preston R.R., Vil M.D., Baker J.P., Miller B., Zutavern T.,
RA   Rodriguez S., Santos L., Kuit K.H., Cunnius D.M., Balija V.S.,
RA   Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA   McCombie W.R.;
RT   "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT   Clone OSJNBa0058E19, complete sequence.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC080019; AAK13116.1; -.
DR   InterPro; IPR001995; Aspprotease_rtrv.
DR   InterPro; IPR001969; Aspprotease_site.
DR   InterPro; IPR005162; Retrotrans_gag.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF03732; Retrotrans_gag; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; rvp; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW   RNA-directed DNA polymerase.
SQ   SEQUENCE   2162 AA;  244771 MW;  826F55524A5C67EF CRC64;

  Query Match             81.8%;  Score 36;  DB 10;  Length 2162;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TAYRYH 6
          ||||||
Db    483 TAYRYH 488


RESULT 14
Q9HF73
ID   Q9HF73      PRELIMINARY;      PRT;  211 AA.
AC   Q9HF73;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Yer153p (Fragment).
OS   Ashbya gossypii (Eremothecium gossypii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Eremotheciaceae; Eremothecium.
OX   NCBI_TaxID=33169;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wendland J., Springer N., Philippsen P.;
RT   "BEM2 is required for polarized growth and maintenance of cell
RT   polarity in the filamentous ascomycete Ashbya gossypii.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF195007; AAG43465.1; -.
FT   NON_TER     211     211
SQ   SEQUENCE   211 AA;  24538 MW;  74EBF96DC8CE7A1E CRC64;

  Query Match             79.5%;  Score 35;  DB 3;  Length 211;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 YRYHLL 8
          ||||||
Db    118 YRYHLL 123


RESULT 15
Q8UAD4
ID   Q8UAD4      PRELIMINARY;      PRT;  323 AA.
AC   Q8UAD4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional regulator, LysR family.
GN   ATU3440 OR AGR_L_2774.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009274; AAL44253.1; -.
DR   EMBL; AE008338; AAK89952.1; -.
KW   Complete proteome.
SQ   SEQUENCE   323 AA;  35071 MW;  BEDBF296CBFFD9F4 CRC64;
```

Query Match 79.5%; Score 35; DB 16; Length 323;
Best Local Similarity 87.5%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 TAYRYHLL 8
             || ||||
Db     287 TASRYHLL 294
```

Search completed: May 22, 2003, 12:09:40
Job time : 28 secs